STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF AN INTRAMEMBRANE PEPTIDASE AND A NON-PEPTIDASE HOMOLOG

A Dissertation
Presented to
The Academic Faculty

by

Sibel Kalyoncu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2016

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF AN INTRAMEMBRANE PEPTIDASE AND A NON-PEPTIDASE HOMOLOG

Approved by:

Dr. Raquel Lieberman, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Ingeborg Schmidt-Krey
School of Biology
Georgia Institute of Technology

Dr. Loren Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Xiaodong Cheng
Department of Biochemistry
Emory University

Dr. Bridgette Barry
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: June 9, 2016

To Erdal and our baby to be born. I am so lucky to have you in my life.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Raquel Lieberman who supported and guided me during all those years. Her valuable mentorship made me a better scientist. She will keep being my role model throughout my future career.

Next, I would like to thank all current and past members of Lieberman Lab. Special thanks to David Heaner (an amazing undergraduate who worked with me for 4 years) for his invaluable contribution to my research, Dr. Jennifer Johnson for her help during my first semester, and Swe-Htet Naing for his help and smiles brought to the lab. I feel lucky to have the opportunity to work with Dr. Rebecca Donegan, Dr. Shannon Hill, Dr. Dustin Huard, Elaine Nguyen, Athéna Patterson-Orazem, Chiamaka Ukachukwu, Casey Bethel (Teacher of the year!), Shuo Huang, Dr. Jason Drury, Jeff Culver, Katherine Turnage, Dr. Susan Orwig, Dr. Tanay Dasai, Lindsay Porter, Michelle Womack-Davis. Lieberman lab would not be complete without all fantastic undergraduates that we have had, thanks to James Going, Ivan Morales, James Rives, Aly Sheppard, Michelle Kwon, Josh George, Quincy Faber, Dana Freeman.

I would like to thank my committee members, Dr. Loren Williams, Dr. Inga Schmidt-Krey, Dr. Bridgette Barry and Dr. Xiaodong Cheng for their guidance and input to this dissertation. I have collaborated with many labs across the country and I would like to thank for their contribution to my research. Special thanks to Dr. Jennifer Maynard and her lab (Jeongmin Hyun, Dr. Kevin Entzminger) for their work on antibody fragments, Dr. Jim Spain and his lab (Dr. Zohre Kurt) for their fruitful collaboration of

the most amazing enzyme, Dr. Schmidt-Krey and her lab (Maureen Metcalfe and Kasahun Neselu) for their collaboration of 2D crystallization.

Finally, I would like to thank my beloved family. Special thanks to my husband, Dr. Erdal Uzunlar, for his never-ending love and encouragement, and the beautiful family with a sweet baby to be born. I would not have reached this goal without you. I want to thank to my mom, dad, siblings and in-laws for their support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

[S]	Substrate concentration
°	Degrees
° C	Degrees Celsius
2D	2 Dimensional
3D	3 Dimensional
3D5/EE	anti-EE scFv derived from 3D5
3D5/His	anti-His_6 scFv derived from 3D5
4NA	4-Nitroaniline
5NAA	5-Nitroantranilic acid
5NAA-A	5-Nitroantranilic acid aminohydrolase
5NSA	5-Nitrosalicylic acid
Å	Angstrom
AA	Anthranilic acid
ACD	N-acetyl-L-citrulline deacetylase
AcrB	Acriflavine resistance protein from *E. coli*
ACS	American Chemical Society
AD	Alzheimer's disease
AFU	Arbitrary fluorescence unit
Anapoe-C12E8	Polyoxyethylene(8)dodecyl Ether
ANB	4-Amino-3-nitrobenzoic acid
Anti-EE	Specific to EE

Anti-His_6	Specific to His_6
Apo	Protein without its cofactor(s)
APP	Amyloid precursor protein
ATPA	Aminoterephthalic acid
Aβ	Amyloid beta peptide
BioCAT	Biophysics Collaborative Access Team
BLAST	Basic Local Alignment Search Tool
BNBA	2-Bromo-5-nitrobenzoic acid
c	Protein concentration
CD	Circular dichroism
CDR	Complementarity determining region
CHAPSO	3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate
CMC	Critical micelle concentration
CPM	*N*-[4- (7-Diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide
Cryo-EM	Cryo-electron microscopy
CV	Column volume
d	Path length
Da	Dalton
DARPins	Designed ankyrin repeat proteins
DDM	n-dodecyl-β-D-maltopyranoside
DM	n-Decyl-β-D -maltopyranoside
D_{max}	Maximum inter-atomic distance

DMPC	1,2-Dimyristoyl-sn-glycero-3-phosphocholine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAN	2,4-Dinitroaniline
EBB	Engineered biosystems building
E. coli	Escherichia coli
E.C.	Enzyme classification
EDTA	Ethylenediaminetetraacetic acid
EE	Amino acid sequence EYMPME
ER	Endoplasmic reticulum
ES	MCM-EE with shortened loop
ETS	MCM-EE with shortened loop and TEV cleavage site before C-terminal His_6 tag
Fab	Fab antibody fragment
FAD	Familial Alzheimer's disease
FC-12	Fos-choline 12
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
FSEC	Fluorescent size exclusion chromatography
GFP	Green fluorescent protein
GlpG	Rhomboid Protease
GPCR	G protein-coupled receptor
$H_2{}^{18}O$	Water with ^{18}O isotope

HBS	HEPES buffered saline
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	Hexahistidine (amino acid sequence HHHHHH)
HIV	Human immunodeficiency virus
HLA-E	Histocompatibility antigen, alpha chain E
Holo	Protein with its cofactor(s)
hSPP	SPP from *Halobacterium salinarum*
hSPP-EE	EE-tagged SPP from *Halobacterium salinarum*
HsSPP	SPP from human
I(q)	Scattered intensity
IAP	Intramembrane aspartyl protease
IC_{50}	Half maximal inhibitory concentration
Intimin-EE	EE-tagged E. coli intimin
IP	Intramembrane peptidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_a	Association constant
k_{cat}	Turnover number
K_D (or K_d)	Dissociation constant
kDa (or K)	Kilodalton
keV	Kiloelectron volt
kJ	Kilojoules
K_m	Substrate concentration at half maximal velocity

k_{off}	Dissociation rate constant
k_{on}	Association rate constant
kV	Kilovolt
L	Liter
LB	Luria-Burtani
LC-MS/MS	Liquid chromatography–tandem mass spectrometry
LCP	Lipidic cubic phase
LDAO	N,N-Dimethyldodecylamine-N-oxide
LFD	Lipopeptide detergents
LIC	Ligation-independent cloning
LMNG	Lauryl maltose neopentyl glycol
M	Molar
mA	Miliamper
MBP	Maltose binding protein
MCM-EE	EE-tagged SPP from *Methanoculleus marisnigri JR1*
MCM-EE-TEV	EE-tagged SPP from *Methanoculleus marisnigri JR1* with TEV cleavage site before C-terminal His_6 tag
MCMSPP	SPP from *Methanoculleus marisnigri JR1*
MDa	Megadalton
MEROPS	A peptidase database
MHC	Major histocompatibility complex
MmRce1	*Methanococcus maripaludis* Rce1
mNBA	m-Nitrobenzoic acid

M_{res}	Mean residue mass calculated from the protein sequence
MRS substrate	MBP-IHPFHLVIHT-SUMO-His_6
MS	Mass spectrometry
mSPP	SPP from *Haloarcula marismortui*
mSPP-EE	EE-tagged SPP from *Haloarcula marismortui*
mT4L	Modified T4 lysozyme fusion protein
MWCO	Molecular weight cut-off
NG	n-Nonyl-β-D-Glucopyranoside
NIH	National Institute of Health
NM	n-Nonyl-β-D-Maltopyranoside
NMR	Nuclear magnetic resonance
OD_{600}	Optical density measure at 600 nm
OG	n-Octyl-β-d-glucopyranoside
P(R)	Pair-distance distribution function
PAL	Conserved Proline-Alanine-Leucine motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDBePISA	PDBe Protein Interfaces, Surfaces and Assemblies
PDC	Protein detergent complex
PEG	Polyethylene glycol
pKa	Logarithmic acid dissociation constant
PSH	Archaeal presenilin/SPP homologue

psi	Pound per square inch
q	Scattering vector
R.M.S.D.	Root mean squared deviation
Ren390	IHPFHLVIHT peptide
Ren390FRET	Ren390 with fluorophore and quencher
RF	Restriction free
R_g	Radius of gyration
RIP	Regulated intramembrane proteolysis
R_{max}	Maximum analyte binding capacity in RU
RPM	Revolutions per minute
RU	Response units
S12	Superose 12 10/300 GL
S2P	Site-2 protease
S300	HiPrep 16/60 Sephacryl S-300
scFv	Single-chain antibody fragment or single-chain variable fragment
ScNP	*Streptomyces caespitosus* neutral protease
SDAD	N-succinyl-L,L-diaminopimelic acid desuccinylase
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-SAXS	Size-exclusion chromatography - small angle x-ray scattering
SeMet	Selenomethionine
SER	Surface entropy reduction

SER-CAT	Southeast Regional Collaborative Access Team
S_NAR	Nucleophilic aromatic substitution
SPP	Signal peptide peptidase
SPR	Surface plasmon resonance
SUMO	Small ubiquitin-like modifier protein
T4L	T4 lysozyme
TB	Terrific broth
Tet	(R)-1-carboxy-2-amino-2-hydroxy-5-nitro-cyclohexa-1,4-diene
TEV	Tobacco etch virus
TLC	Thin layer chromatography
T_m	Melting temperature
TM	Transmembrane helix
TXRF	Total reflection x-ray fluorescence
UT	University of Texas
UV-Vis	Ultraviolet–visible
V	Initial velocity
V_H	Antibody variable heavy chain
V_L	Antibody variable light chain
V_{max}	Maximum velocity at infinite substrate concentration
w/v	Weight/volume
WT	Wild-type
ZLL_2-ketone	An SPP-specific inhibitor
ZMPSTE24	Zinc metallopeptidase STE24

ΔG	Change in the Gibbs free energy
$\Delta H°$	Change in enthalpy
$\Delta N23$	Without N-terminal 23 amino acids
λ_{ex}	Excitation wavelength
λ_{em}	Emission wavelength
λ_{max}	Maximum wavelength
μ	Micro
Φ	Ellipticity

SUMMARY

Peptidases play fundamental roles in all living organisms and their dysfunction is associated with a variety of diseases. Although sequences of peptidases encoded in genomes throughout life have become readily available via high throughput sequencing technologies, research on their structural and functional characterizations lags behind due to challenges related to their crystallization and time-intensive biochemical/biophysical studies.

Signal Peptide Peptidase (SPP) is an intramembrane aspartyl peptidase that cleaves signal peptides within the hydrophobic region of cellular membrane. SPP plays important roles in cellular functions such as immune system regulation. Structural characterization of membrane proteins including SPP is challenging due to their hydrophobic content which is undesirable in crystallization. Structures of membrane proteins are severely underrepresented: number of unique membrane protein structures is still less than 1% in Protein Data Bank. Here, a new generalizable method was introduced to overcome crystallization challenge of membrane proteins. A toolbox of single chain antibody fragments (scFvs) specific to the EYMPME peptide (EE) epitope was developed for use as co-crystallization chaperones. Structures of designed scFvs were solved and their crystallization propensities were systematically explored to improve their chaperone abilities. Tight complexation of anti-EE scFvs with an EE-tagged archeal SPP and another test membrane protein was demonstrated. Important lessons learned during crystallization and co-crystallization trials of scFvs and SPP are discussed in this dissertation.

To understand peptidases at a mechanistic level requires both high resolution structures and extensive structure-function studies in which residues are systematically altered and differences in functionality of the peptidase are measured. Although a low resolution structure of inactive SPP became available during my PhD studies, details on how SPP recognizes and catalyzes its substrate are still not known. Here, preliminary data for a structure-function study to understand substrate gating mechanism of SPP are presented.

Finally, structure-function studies of 5-nitroanthralinic acid aminohydrolase (5NAA-A), a metallo-peptidase family member that catalyzes a deamination reaction on a natural, toxic nitroaromatic compound, are presented. 5NAA-A has evolved a function other than peptide hydrolysis but is structurally and evolutionary related to peptidases and is thus classified as a non-peptidase homolog. We characterized 5NAA-A biochemically and biophysically, and obtained snapshots of its mechanism by solving its crystal structures in various states. The 5NAA-A structure and its nucleophilic aromatic substitution mechanism expand our understanding of the great diversity of enzymes capable of transforming natural organic compounds in our ecosystem.

CHAPTER 1: Introduction to Peptidases

1.1. Roles and function

1.1.1. Hydrolysis of peptide bonds

Peptidases (proteases, proteinases, proteolytic enzymes) perform peptide bond hydrolysis (**Figure 1.1**). Peptidases are one of the most abundant families of enzymes, accounting for 1-2% of all protein coding genes in a genome (*1*), and are present in all cellular organisms. Peptidases and their homologs fall into different families corresponding to distinct catalytic mechanisms and substrate specificities (Section 1.2).

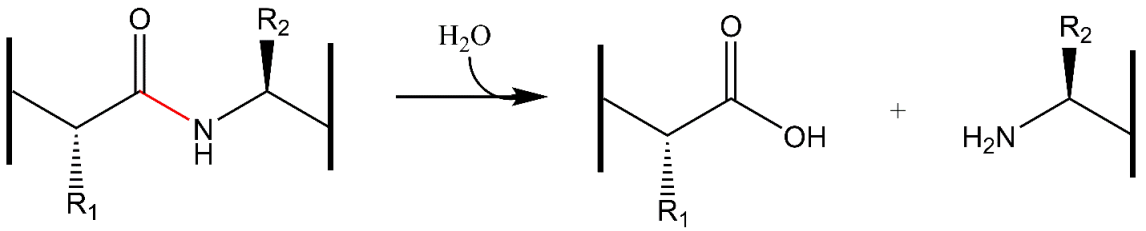


Figure 1.1. Peptide bond hydrolysis. Scissile peptide bond is colored in red.

The majority of peptidases catalyze the hydrolysis of α-peptide bond(s) between natural amino acids (**Figure 1.1**) but some peptidases perform reactions that are slightly different or have additional features. For example, ubiquitin carboxyl-terminal hydrolase 1 can catalyze an isopeptide bond in an ubiquitin conjugate (*2*), glutamate carboxypeptidase II can hydrolyze a γ-glutamyl bond (*3*), and γ-glutamyl transpeptidase 1 can both transfer and hydrolyze peptide bonds (*4*).

1.1.2. Functions of peptidases in biological processes

Peptidases are involved in a wide variety of biological processes, including cell cycle/proliferation/survival/death, signaling pathways, protein trafficking and immune response. Peptidases thus also play important roles in many human diseases such as cancer, inflammatory, infectious, cardiovascular and degenerative diseases (*5, 6*) and are attractive targets for the pharmaceutical industry due to their potential as diagnostic/prognostic biomarkers and drug targets (*7*). Currently, approved protease inhibitor drugs on the market treat hypertension, coagulation disorders, HIV, diabetes and cancer (*5, 8*). In addition to biological/therapeutic impacts, there are broader applications of commercial peptidases such as food processing, detergent and leather industry (*9*). In this chapter, peptidases will be reviewed from a biological and biochemical/biophysical standpoint.

Early studies of peptidases focused on their function in protein catabolism, namely, degrading/recycling intracellular and extracellular proteins as a source of amino acids for the cell. However, peptidases play many other key roles in which highly specific reactions are required to control a biological process (*10*). Peptidases mediate post-translational modifications such as signal peptide removal and protein activation by pro-peptide removal. In this capacity, peptidases can regulate the localization and activity of other cellular proteins as well as control protein-protein interactions. Localization of peptidases and their substrates control their response to different stimuli (*11*). Peptidases can also deactivate certain proteins, including peptide hormones. Finally, an active area of research is the roles of peptidases playing in signaling pathways (*12, 13*).

1.2. Classification, mechanism, structure and regulation

1.2.1. Classification based on catalytic mechanism

Traditionally, peptidases were classified into two main groups: (i) endopeptidases, which catalyze internal peptide bonds, and (ii) exopeptidases, which target amino or carboxy terminus of a protein. However, accumulation of structural and biochemical information has led new classifications. The standard classes of peptidases are based on their catalytic mechanisms, of which there are six identified by their catalytic nucleophile (*1, 13, 14*). Serine, cysteine, threonine classes use their namesake residue as the nucleophile, which forms the acyl-enzyme intermediate (**Figure 1.2A-C**). For example, serine peptidases use a catalytic triad composed of Ser-His-Asp (**Figure 1.2A**) (*15*) where the nucleophilicity of Ser is enhanced by His and Asp to form a tetrahedral intermediate stabilized by the so-called oxyanion hole formed by additional residues in the binding pocket (Gly and another Ser). Aspartate, glutamate and metallo- peptidases use an activated water as a nucleophile for acid-base catalysis (**Figure 1.2D-F**). The water is activated through a carboxylic acid group of the catalytic residue or a metal ion. In metallo-peptidases, a divalent cation, often zinc, is used to reduce the pK_a of the water molecule for further activation. The divalent metal is coordinated by at least three active site residues (usually His, Glu and Asp) and the water molecule. Some metallo-peptidases, such as carboxypeptidase G2, are di-nuclear with two metal ions in the active site (*16*). A seventh class, asparagine peptide lyases, which self-cleave at asparagine residues (*17*), does not involve conventional hydrolysis; thus, its inclusion in the peptidase classification remains unresolved.

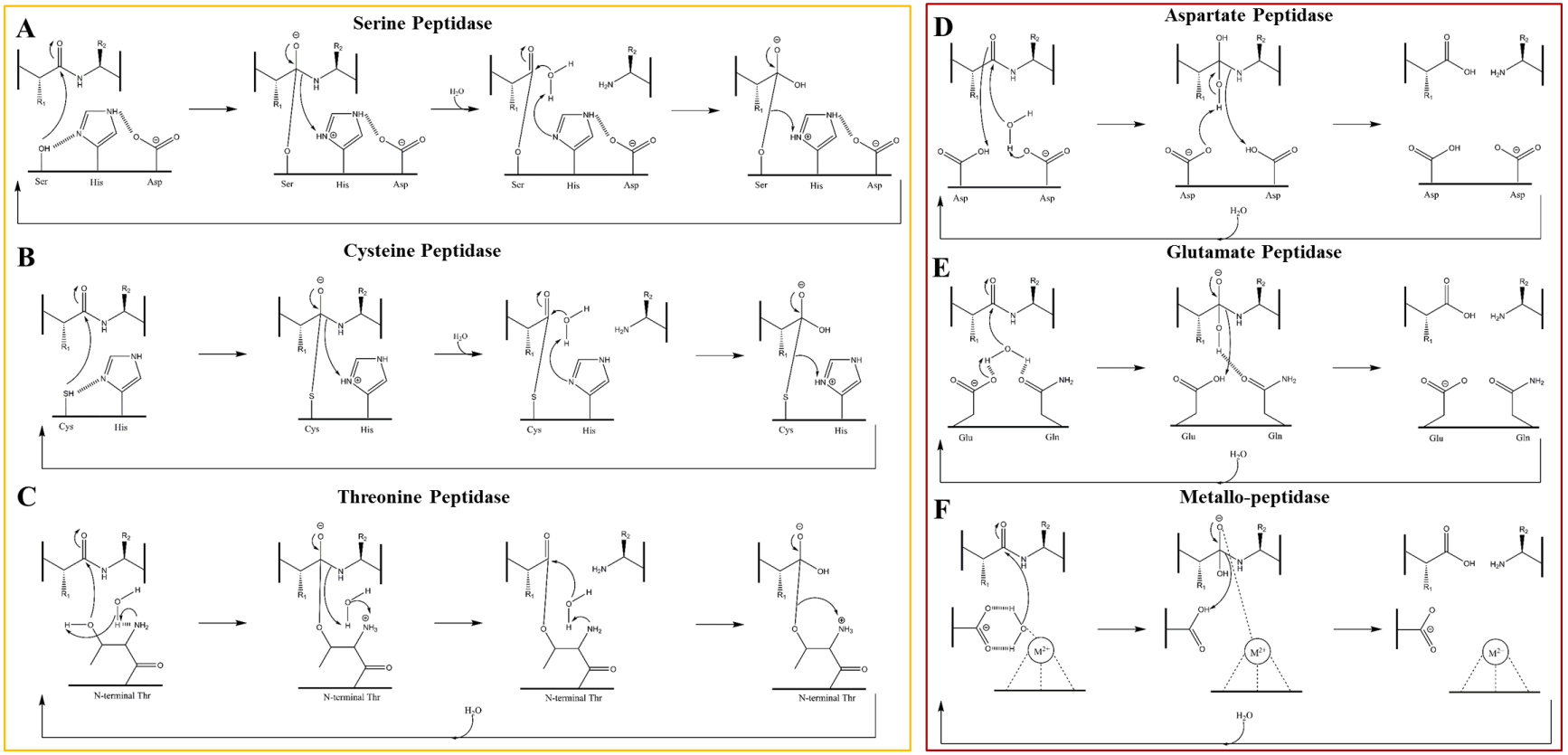


Figure 1.2. Proposed catalytic mechanisms of six peptidase classes. **(A-C)** Mechanisms of serine, cysteine and threonine peptidases respectively. An acyl-enzyme intermediate forms between the carbonyl carbon of the peptide substrate and the corresponding catalytic residue. Threonine peptidase uses an N-terminal Thr for nucleophilic attack because N-term acts as a general base. **(D-F)** Mechanisms of aspartate, glutamate and metallo- peptidases respectively. A water molecule is activated for a nucleophilic attack on the carbonyl carbon of the peptide substrate.

1.2.2. Structural features

Although they perform parallel mechanisms (**Figure 1.2**), structures vary widely among different classes of peptidases. Among the six peptidase classes, the most diverse is the metallo-peptidases (*18*). Some peptidases are as small as 15 kDa, with only a small catalytic domain like the bacterial zinc peptidase ScNP (*19*) and some are as large as 1-6 MDa with many domains like the multimeric meprin metalloproteases (*20*) or the proteasome (*21*). Many peptidases link their catalytic domain to specialized domains to tailor their cellular localization, substrate specificity, kinetic properties and/or inhibitor sensitivity. The level of complexity in peptidase structure could be the driving force for the evolution of primitive catabolic enzymes to highly specialized enzymes that function in cellular processes.

1.2.3. Specificity and promiscuity

Peptidases bind to their substrates through a combination of hydrogen bonding, electrostatic, and hydrophobic interactions with substrate peptide backbones and/or side chains. While some peptidases show high level of specificity targeting unique peptide bond such as angiotensin-converting enzyme (*22*), many are relatively non-specific and promiscuous, such as proteinase K (*23*). The identification of peptidase substrates and analysis of their specificity towards those substrates has become an important focus of current research (*24*). Two main approaches have been commonly used for this purpose: (i) gel-based methods where complex mixtures are separated and identified by mass spectrometry (MS) analysis, (ii) N- and C-termini based proteomics strategies where cleaved substrate is enriched for MS analysis (*25, 26*). To gain functional information on

large numbers of protein, library-based approaches have been developed such as mixture-based peptide libraries (*27*) and multiplex substrate profiling (*28*).

1.2.4. Regulation and activation

Peptidases are highly regulated in the cell because they catalyze essentially irreversible reactions and must respond to specific cellular cues temporally and spatially. The known regulation mechanisms for peptidases include gene expression, post-translational modifications, zymogen activation via limited proteolysis, cellular localization, co-factors, allosteric mechanisms, endogenous peptidase inhibitors and degradation (*13, 29, 30*). Protein reservoir of the encoded genome can be extensively diversified by regulated peptidase mechanisms. How peptidase gene expression is controlled is generally not known, but notable exceptions include matrix metalloproteases that are implicated in cancer (*31*).

Inactive zymogens can be activated by either autocatalysis or catalysis via a different peptidase. For example, presenilin performs autocatalysis to become active and process amyloid precursor protein which plays important role in Alzheimer's disease (*32, 33*). Blood coagulation cascade involves sequential serine protease activation (*34*). Dysregulation of peptidases can lead to cell death; apoptosis can occur when pro-growth cell cycle proteins are not properly degraded (*35*).

1.3. Intramembrane peptidases

Intramembrane peptidases (IPs), of particular interest in this thesis, act within the hydrophobic cell membrane to cleave transmembrane helices. Intramembrane proteolysis was first observed less than three decades ago (*36*). Soluble and intramembrane

peptidases are similar in their requirement of substrate recognition before processing. However, data to date indicate that compared to soluble peptidases which cleave their substrates within seconds, IPs cleave substrates on a timescale of minutes (*37, 38*). The details of IP biochemistry are limited but their malfunction is associated with pathologies such as Alzheimer's and Parkinson's diseases (*39, 40*). More details on classes of IPs and their proposed mechanisms are presented in Chapter 2.

Although the structures of some IPs have been solved and increasing number of studies have been conducted to understand their mechanism of action (*41-44*), there are still many open questions: (i) how do IPs recognize their substrates and distinguish them from other transmembrane helices, (ii) what is the rate limiting step responsible for the slow kinetics, (iii) what kind of conformational changes occurs on enzyme and/or substrate for substrate recognition and processing, (iv) how do substrates gain access to the hydrophilic catalytic site, (v) how do disease-related mutations affect the mechanism, (vi) how does the membrane environment such as lipids affect the cleavage.

1.4. Non-peptidase homologs

1.4.1. Similar structures but different chemical reactions

Peptidases are classified based on their evolutionary and structural relationships (*18*), not explicitly by the reactions they catalyze (Section 1.2.1). For example, MEROPS, a peptidase database regularly updated and very useful for many researchers worldwide (*45*), clusters homologous sequences expected to share a similar structure and/or catalytic type. Due to advances in genome sequencing, functional characterization of hypothetical proteins encoding predicted peptidases lags behind the availability of their corresponding translated nucleotide sequences and phylogenetic analysis (*46*). By noting the

replacement of key active site residues, some uncharacterized enzymes can be annotated as non-peptidases homologs (*47*), but because it is not possible to identify substrates from primary sequence or even atomic structure, there exist sequences within peptidase families that perform chemistry other than peptide bond hydrolysis.

1.4.2. An example: 5-nitroantranilic acid aminohydrolase (5NAA-A)

5NAA-A is a metallopeptidase family member that has evolved to hydrolyze a toxic nitroaromatic compound, 5-nitroantranilic acid (5NAA), to 5-nitrosalicylic acid (5NSA) (*48*). As discussed in Chapter 6, we characterized 5NAA-A biochemically/biophysically and solved its structure in various states (*49*). Its sequence, structure, and biochemical characterization confirm that 5NAA-A belongs to the MEROPS M20 metallopeptidase family as a non-peptidase homolog (*50*) and evolved from a common ancestor (*51*). The closest structural homologues are other M20 family members carboxypeptidase G2 (root mean squared deviation (R.M.S.D.): 5.04 Å, (sequence similarity ~35%) (*16*), N-acetyl-L-citrulline deacetylase (ACD, R.M.S.D.: 4.32 Å) (*52*), and N-succinyl-L,L-diaminopimelic acid desuccinylase (SDAD, R.M.S.D.: 5.66 Å) (*53*) (**Figure 1.3**). However, 5NAA-A harbors key distinctions leading to its unique supramolecular arrangement, catalytic site, and active site chemistry (Chapter 6).

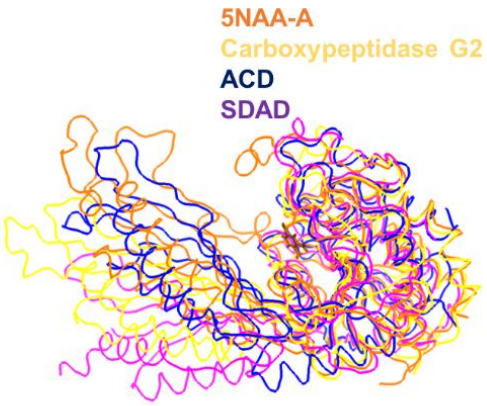


Figure 1.3. Monomer superposition of MEROPS M20 family members: 5NAA-A (orange), carboxypeptidase G2 (yellow, PDB ID: 1CG2), N-acetyl-L-citrulline deacetylase (blue, ACD, PDB ID: 2F7V), and N-succinyl-L,L-diaminopimelic acid desuccinylase (purple, SDAD, PDB ID: 3IC1).

By phylogenetic analysis, we identified additional MEROPS M20 non-peptidase homologs that likely harbor mononuclear metal binding sites and hydrolyze other aromatic substrates of potential environmental or synthetic interest (Chapter 6). It appears likely that many additional rare and unusual non-peptidase homologs are involved in catabolic pathways for the great diversity of natural organic compounds in our ecosystem (*54*).

1.5. Thesis objectives

1.5.1. Development of single chain antibody fragments (scFvs) for membrane protein co-crystallization

Structural and biochemical/biophysical characterization of membrane proteins is challenging due to their hydrophobic content. In chapter 2, I present methods used to prepare an archaeal signal peptide peptidase (SPP), ortholog of an intramembrane

peptidase in humans that plays essential roles in immune response and surveillance. Several optimization strategies ranging from molecular cloning to purification to detergents were employed to obtain high yield, pure and active SPP for further characterization.

The next objective was to develop antibody fragments specific to a short peptide epitope that can be easily installed in an extramembranous loops of any membrane protein target for facile complexation. These antibody fragments would then aid crystallization of membrane proteins by increasing their hydrophilic crystal contact interfaces. As presented in chapter 3, several single chain antibody fragments (scFvs) were developed from a parent anti-His_6 scFv, in collaboration with the lab of Jennifer Maynard at University of Texas (UT), Austin. Biophysical properties of parent scFv were improved and specificity was then converted to anti-EYMPME (EE). After solving the structures of scFvs, rational design was performed to manipulate the crystallization lattice to better accommodate a membrane protein of interest. Lessons regarding the challenges of controlling crystal lattice contacts were learned from this study.

In chapter 4, I present results of two test membrane proteins targeted for anti-EE scFv complexation and co-crystallization trials: SPP as an α-helical membrane protein and *E. coli* intimin as a β-barrel membrane protein. Only one site-directed-mutagenesis step was required to install EE-tag in selected soluble loops. Several parameters were optimized including the location of EE-tag and solution complexation to form a tight interaction between the anti-EE scFvs and EE-tagged membrane proteins for co-crystallization trials.

1.5.2. Protein engineering and several optimizations for SPP to improve its crystallization and diffraction

In chapter 5, I present my efforts to improve crystallization and diffraction quality of SPP crystals. This chapter catalogs the exhaustive efforts to crystallize SPP by itself. SPP crystallization was optimized via alterations in detergent/lipid environment, by protein engineering, and several other commonly used techniques in the field. Unfortunately, although diffraction quality of the crystals was significantly improved by my efforts, I determined that the protein crystals were not of SPP but of a contaminant. Thus, while I did not solve a high resolution structure of SPP, I learned valuable lessons regarding contamination issues for any membrane protein crystallization experiments.

1.5.3. An on-going structure-function study to understand substrate gating mechanism of SPP

During the course of this thesis work, a 3.8 Å resolution crystal structure of SPP in an inactive conformation became available in the literature. Using this structure, 7 mutants were designed to probe substrate gating in SPP, taking advantage of the lab's expertise in enzyme kinetics and biophysical characterization. I developed a protocol to measure substrate binding on-off kinetics by surface plasmon resonance (SPR) to differentiate between mutants that reduce activity via impaired substrate binding and those for which binding is similar to that of wild-type SPP. Preliminary results are presented in chapter 7, which will be repeated and expanded by Swe-Htet Naing, another PhD candidate in Lieberman Lab, after my departure. Chapter 7 thus provides a perspective on future directions for the project.

1.5.4. Structural and biochemical/biophysical characterization of a novel metalloenzyme, 5NAA-A, capable of hydrolyzing a toxic, natural nitroaromatic compound

Chapter 6 presents results from a long-standing collaborative side project between the Lieberman lab and that of Jim Spain from the School of Civil and Environmental Engineering at Georgia Tech. The goal of this project was to structurally and biochemically/biophysically characterize a novel metalloenzyme, 5NAA-A, classified among metallopeptidases but instead hydrolyzes a toxic natural nitroaromatic compound, 5NAA to 5NSA. We obtained crystallographic snapshots of the mechanism and explored its substrate specificity.

CHAPTER 2: Molecular Cloning, Expression, Purification and Detergent Optimization of Signal Peptide Peptidase (SPP)

2.1. Introduction

2.1.1. Intramembrane peptidases (IPs)

Regulated intramembrane proteolysis (RIP) plays various roles in cell signaling events that are important for cell proliferation, differentiation, development, and metabolism (*55*). Peptidases involved in RIP are called intramembrane peptidases (IPs). IPs cleave within transmembrane region of membrane-bound substrates to release cytoplasmic and/or extracellular peptides, which translocate to different parts of the cell to perform their corresponding biological response (*56*). There are four classes of IPs identified to date based on their catalytic units (*55-57*): (i) serine peptidases called rhomboids which play roles in the epidermal growth factor receptor pathway (*58*), (ii) zinc metallopeptidases called site-2-proteases (S2P), which are involved in the regulation of cholesterol metabolism (*59*), (iii) glutamate proteases which involve in cell differentiation and carcinogenesis (*41*), and (iv) aspartyl proteases (*42*) which will be discussed in detail in section 2.1.2. These different classes of IPs are not evolutionary related but they are all polytopic transmembrane α-helical proteins (**Figure 2.1**). Despite their essential functions in the cell, details of their biochemistry, mechanism and substrate specificity are not well understood yet.

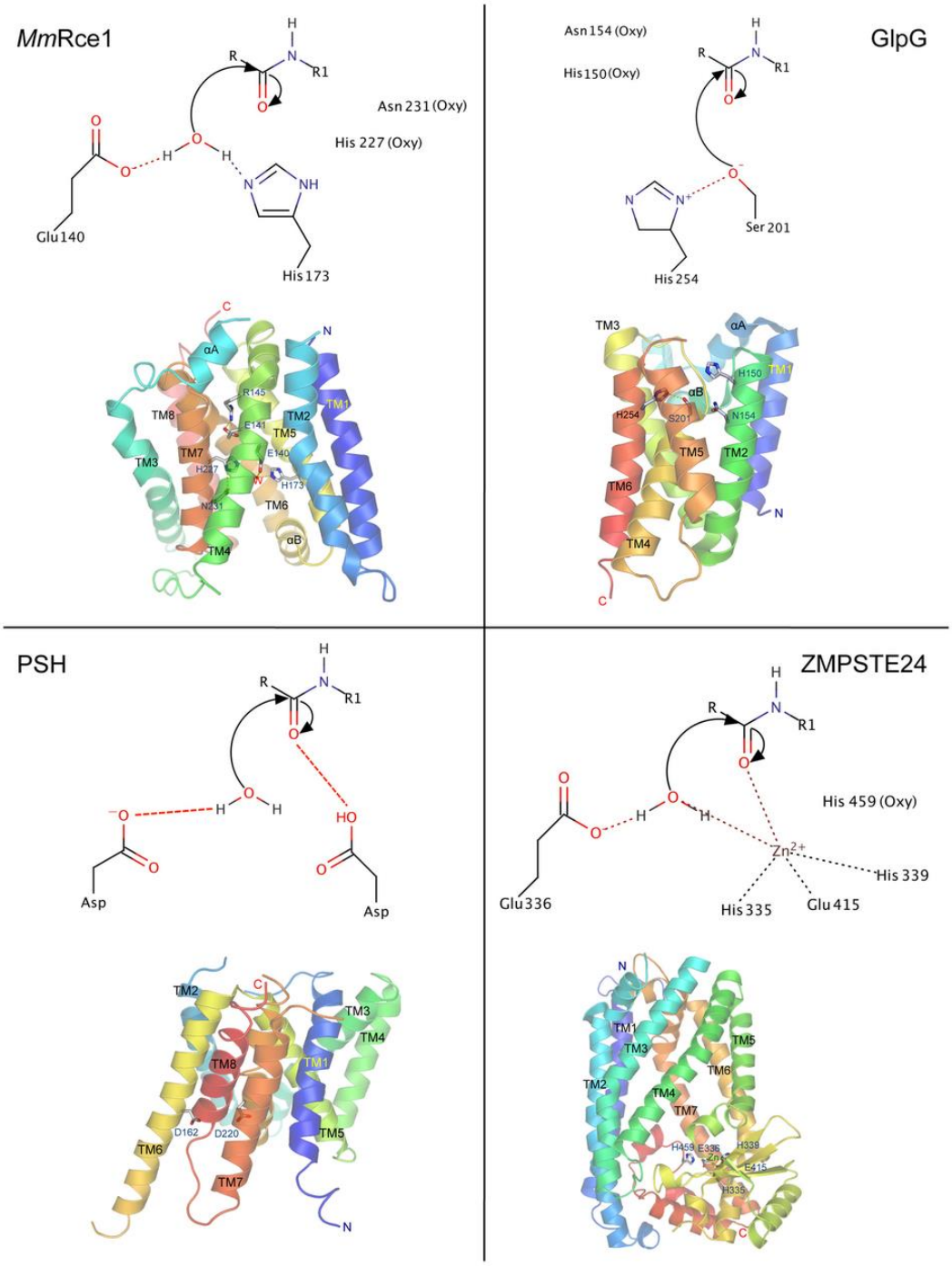


Figure 2.1. The proposed catalytic mechanisms and overall structures of intramembrane proteases (IPs). Glutamate IP (MmRce1, PDB ID 4CAD) (*41*), serine IP (GlpG, PDB ID 2O7L) (*58*), aspartyl IP (PSH, PDB ID 4HYC) (*42*) and zinc metalloprotease S2P (ZMPSTE24, PDB ID 4AW6) (*59*). Structures are viewed parallel to the membrane, oriented from the ER lumen (top) to the cytoplasm (bottom). Reprinted by permission from Macmillan Publishers Ltd (*41*).

2.1.2. Intramembrane aspartyl proteases and SPP

Intramembrane aspartyl proteases (IAPs) includes presenilin and Signal Peptide Peptidase (SPP) (*33, 60*). It has been shown that mutations in presenilin are related to early onset Alzheimer's disease (AD) (*32*). Presenilin is the catalytic subunit of γ-secretase, it sequentially cleaves amyloid precursor protein (APP) to generate different amyloid β (Aβ) peptides that is deposited in the brain of AD patients. Presenilin is very similar to SPP in structure, with 9 transmembrane helices but opposite topology (**Figure 2.2**). Their active sites include the same conserved motifs ("YD" and "GXGD") and it was shown that they are inhibited by the same active-site directed compounds (*61*). Unlike SPP, presenilin performs self-cleavage to become catalytically active (*57*).

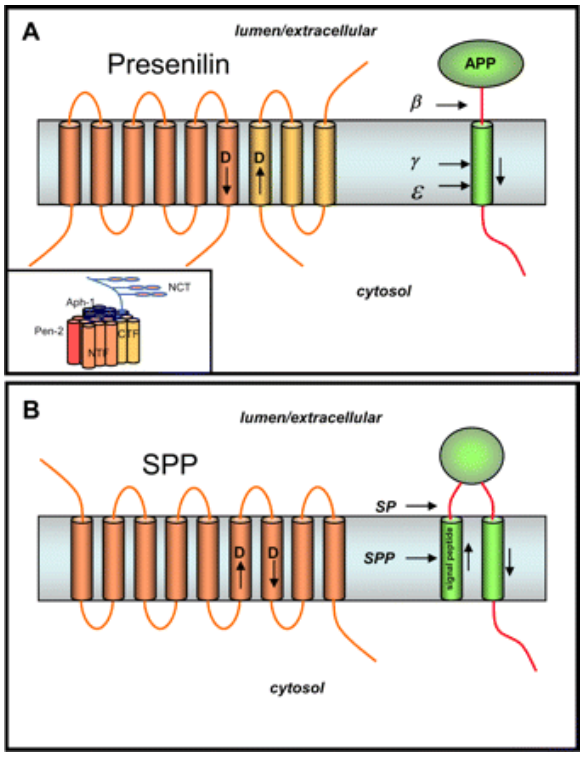


Figure 2.2. Intramembrane aspartyl proteases. (**A**) Presenilin is the catalytic component of γ-secretase complex (inset). It cleaves APP to form Aβ peptides. (**B**) SPP processes signal peptides following signal peptidase (SP) cleavage. Reprinted by permission from American Society for Biochemistry and Molecular Biology (*33*).

SPP is an IAP with orthologs found from human to archaea (*62*). In eukaryotes, after signal peptidase cleaves the signal peptide of a nascent protein, SPP cleaves type-2 signal peptides from the endoplasmic reticulum (ER) membrane. SPP plays key roles in immune surveillance; it processes peptides such as those from newly synthesized major histocompatibility complex (MHC) class I molecules (*63*) to form histocompatibility antigen HLA-E epitopes which are presented on the cell surface for signaling natural killer cells that protein synthesis is proceeding normally (*64*). Also, SPP is known to process N-terminal domain of hepatitis C virus for virus maturation (*65*).

2.1.3. Chapter overview and publications

Detailed biochemical and structural investigation of membrane proteins are challenging compared to soluble proteins due to their lower protein expression levels and need for detergent and/or lipids in the downstream processes (*66*). Because of its important role in the cleavage of diverse signal peptides and similarity to presenilin, SPP is an attractive target for structural and functional studies. In this chapter, methods applied to optimize SPP expression and purification are discussed.

Two co-author manuscripts were published based on work presented here. First, a book chapter was published in Methods in Molecular Biology volume titled Heterologous Expression of Membrane Proteins: Methods and Protocols, 2nd edition (*67*). Second, a manuscript about a novel SPP peptide cleavage assay was published in ACS Chemical Biology in 2015 (*68*).

2.2. Methods

2.2.1. Molecular cloning of archeal SPP orthologs

Several archeal SPP orthologs were cloned by Dr. Raquel Lieberman during her post-doctoral studies. Among those constructs, *Haloarcula marismortui* SPP (mSPP) and *Halobacterium salinarum* SPP (hSPP) were chosen for further studies. *Methanoculleus marisnigri JR1* SPP (MCMSPP) was cloned separately. Genomes were obtained from ATCC (http://www.atcc.org/). Signal sequence prediction (Signal-3L) (*69*) was performed to detect possible signal sequences, N-terminal 23 amino acids of MCMSPP were predicted as signal sequence. The target SPP sequences were amplified by PCR and cloned into pET-22b(+) vector (Novagen) via restriction sites *Nco*I and *Sal*I (**Figure 2.3**). While cloning of MCMSPP without predicted signal sequence (MCMSPP ΔN23) worked, the one with predicted signal sequence (full-length MCMSPP or MCMSPP) did not work with conventional cloning techniques. Therefore, Jennifer Johnson (former PhD graduate from Lieberman Lab) applied a Restriction Free (RF) cloning method (*70*) to obtain MCMSPP. Construct sequences were confirmed by DNA sequencing (MWG Operon). Primers for MCMSPP cloning are presented in **Table 2.1**.

Figure 2.3. Molecular cloning of archeal SPP orthologs in pET-22b(+) vector via *Nco*I and *Sal*I restriction sites. There is an N-terminal pelB leader sequence for periplasmic membrane insertion and a C-terminal hexahistidine tag for Ni^{2+}-affinity purification (*67*).

Table 2.1. Primers for molecular cloning of MCMSPP. Cloning of full-length MCMSPP only worked with RF cloning.

Constructs	**Forward (F) and Reverse (R) Primers (5'-3')**
MCMSPP	F: GTGGACCATGGGGATGCAGATACGCGACTGGCTGC R: GTACGTGTCGACGAAAGGAAGCCACGAAAACGAACC
MCMSPP (RF cloning)	F: CGCTGCCCAGCCGGCGATGGCCATGCAGATACGCGACTGG R: GCATCGGCATGACGAGGACGATAGCGATGATCTGGACGAAC
MCMSPP ΔN23	F: GTGGACCATGGGGATCGTCCTCGTCATGCCGATGC R: GTACGTGTCGACGAAAGGAAGCCACGAAAACGAACC

2.2.2. Cell growth and protein expression

A standard heat-shock plasmid transformation into *E.coli Rosetta 2 (DE3)* cells were performed and plated onto an agar plate containing 60 µg/mL ampicillin (plasmid resistance) and 34 µg/mL chloramphenicol (cell resistance). After overnight (16-18

hours) incubation at 37°C, a single colony was chosen for expression immediately or the plate was wrapped with a parafilm and stored in the 4°C refrigerator up to one week.

1x200 mL Luria-Bertani (LB, Fisher) broth in a 500 mL Erlenmeyer flask and 6x1-L LB broth in 2 L flasks were autoclaved and 60 µg/mL ampicillin and 34 µg/mL chloramphenicol were added to each flask when cool. A single colony from transformation plate were added into 5 ml LB broth with appropriate antibiotics (starter culture) in the morning and incubated at 37°C for 6-8 hours with shaking at 225 RPM. 200ml LB broth was inoculated with the starter culture and incubated at 37°C for 16-18 hours with shaking at 225 RPM. The next morning, each of 6 1-L cultures were inoculated with 10 mL of the 200 ml culture. For each 1 L culture, starting optical density at 600nm (OD_{600}) should be less than 0.1. 1-L cultures were incubated at 37°C with shaking at 225 RPM, and OD_{600} was checked regularly. When OD_{600} reached 0.6-0.8 (usually takes 3-4 hours), the incubation temperature was reduced to 18°C and shaking was continued for one hour to allow the temperature to equilibrate. Protein expression was induced by the addition of 0.5 mM IPTG (Calbiochem) to each culture and shaking was continued at 225 RPM at 18°C for 16-20 hours.

After expression for 16-20 hours, the cells were harvested by centrifugation at 4200 x g for 10 minutes. Cell pellet was put in a small plastic bag and flash-frozen in liquid nitrogen. They were stored at -80°C freezer until further use.

2.2.3. Membrane isolation

7-8 g of frozen cell paste was resuspended in a 50 mL conical tube on ice by adding 25-30 mL lysis buffer (50 mM HEPES pH 7.5, 200 mM NaCl, Roche complete EDTA Free Protease Inhibitor - 1 tablet to 50 mL buffer). Frequent gentle pipetting was

done with a 25 mL serological pipette until the mixture was fully resuspended and homogeneous. Resuspended cells were lysed using a chilled French press equipment at 1300 psi pressure. The cells were passed through the French press at least twice to ensure complete cell lysis.

Cellular lysate was centrifuged at 5000 x g for 15 minutes at 4°C. Supernatant was transferred in a new centrifuge tube and centrifuged again. This centrifugation step was repeated until no further pellet was discernible (usually 3 times). Final supernatant was transferred to an ultracentrifuge tube and centrifuged at 120,000 x g for 45 minutes at 4°C. The supernatant was discarded and the membrane pellet was placed in a 7 mL Dounce homogenizer with 7 mL lysis buffer (without protease inhibitor). The membrane pellet was resuspended with the loose plunger first for 6 times, then the tight plunger for 10 times. The resuspended membrane pellet was ultracentrifuge again at 120,000 x g for 45 minutes at 4°C. The supernatant was discarded and the pelleted membrane was transferred into a tared microcentrifuge tube in ~0.4 mg aliquots. Its mass was recorded, around 0.9-1.2 g of membrane was typically obtained from 7-8 g lysed cells. Membranes were flash-frozen in liquid nitrogen and stored at -80°C freezer until protein purification.

2.2.4. Protein purification from membranes

Frozen membrane (0.4-0.5 g) was resuspended in 7 mL membrane resuspension buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole) by dounce homogenizer with loose plunger first, then the tight plunger. An amount of the desired detergent (mostly n-dodecyl β-D-maltoside (DDM, Anatrace)) was weighed out equal to the mass of membrane. The detergent was dissolved in enough membrane resuspension buffer to make a final 1% solution. For 0.4 g membrane, 0.4 g detergent was added to 33

mL membrane resuspension buffer, where the last 7 mL would be the resuspended membrane solution. The resuspended membrane was added to the detergent solution and rocked gently at 4°C for 1 hour. Once solubilization is complete the solution appears translucent. Solubilized membrane was centrifuged at 120,000 x g at 4°C for 45 minutes to remove any unsolubilized material. Any remaining pellet was discarded and the supernatant was used for purification.

Protein-detergent solution was loaded into a superloop appropriate for the total volume. With ÄKTA Fast Liquid Chromatography (FPLC) instrument (GE Healthcare), the protein was purified over a 1 mL Ni^{2+}-affinity sepharose chromatography column (GE Healthcare) using Buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.1% DDM) for the wash and Buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole, 0.1% DDM) for protein elution using a gradient elution. The wash step after sample injection was 0% buffer B before but it was changed to 5% buffer B after we discovered a contamination issue (see Chapter 5). The protein amount was monitored by tracking the absorbance at 280 nm. Eluted protein was concentrated to appropriate volume by 50K MWCO Amicon filter for size-exclusion chromatography (SEC) using gel filtration buffer (20 mM HEPES pH 7.5, 250 mM NaCl, 0.05% DDM). HiPrep 16/60 Sephacryl S-300 (S300) or Superose 12 10/300 GL (S12) (GE Healthcare) was used as gel filtration column on an ÄKTA FPLC system. Protein concentration was measured by absorbance at 280 nm using a calculated extinction coefficient. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and coomassie staining was used to assess the purity of the protein.

2.2.5. Detergent exchange by gel filtration

After protein purification by Ni^{2+}-affinity chromatography, the sample was divided into equal volumes, ~250 µL each, to test different detergents by gel filtration. S12 column was equilibrated on an ÄKTA FPLC system with at least two column volumes of gel filtration buffer with corresponding detergents (Anatrace). Each gel filtration buffer contained 2 x the critical micelle concentration (CMC) of corresponding detergent. Used detergents include Fos-Choline-12 (FC-12), N,N-dimethyldodecylamine-N-oxide (LDAO), Lauryl Maltose Neopentyl Glycol (LMNG), Polyoxyethylene(8)dodecyl Ether (Anapoe-C12E8), n-octyl-b-D-glucopyranoside (OG), n-Nonyl-β-D-Glucopyranoside (NG), n-Nonyl-β-D-Maltopyranoside (NM), n-decyl-β-D-maltopyranoside (DM) and n-dodecyl- β-D-maltopyranoside (DDM).

After equilibrating with two column volumes of buffer to remove the prior detergent, the 250 µl sample was injected onto the column. Absorbance at 280 nm was tracked for protein elution. Elution peak shape and intensity was compared. High intensity, Gaussian peaks were desirable because it may show that protein in corresponding detergent is monodisperse. The size of the protein based on elution volume should also be considered, because it may indicate oligomeric state or protein-detergent complex (PDC) size. Further experiments (filter MWCO optimization, activity assay, thermal melt, crystallization, etc.) were conducted with eluted proteins.

2.2.6. Protein stability analysis by Circular Dichroism (CD) melt

Each protein sample was concentrated using a 50K MWCO Amicon filter to 8-10 µM. CD spectra and thermal melt were acquired on a Jasco J-810 spectropolarimeter equipped with Neslab RTE 111 circulating water bath and a Jasco PTC-4245/15

temperature control system. CD spectra were taken at room temperature for both gel filtration buffer (blank) and protein sample in gel filtration buffer. CD thermal melt was performed utilizing a 1 °C/min increase in temperature from 4 to 90°C. Both CD spectra and thermal melt (temperature points for every 2°C) were acquired with 15 averaged scans from 300 to 200 nm at a 200 nm/min scan rate, using a 0.1-cm cuvette.

For each sample, the temperature versus the normalized molar ellipticity was plotted at the minimum wavelength (222 nm for α-helical proteins). Molar ellipticity was calculated from equation: $\Phi = M_{res} \times \Phi_{obs} / 10 \times d \times c$, where M_{res} is the mean residue mass calculated from the protein sequence; Φ_{obs} is the observed ellipticity (degrees) at wavelength 222 nm; d is the path length (cm), and c is the protein concentration (g/ml). The melting temperature (T_m) was determined using molar ellipticity values recorded at 222 nm via sigmoidal dose-response analysis using GraphPad Prism 5 software. High protein stability (high T_m) is desired; detergents that maintain high stability to the protein of interest can be used for further study.

2.2.7. Concentrator Molecular Weight Cut-Off (MWCO) optimization by protein quantification and Thin Layer Chromatography (TLC)

For each purified protein in each detergent to be tested, gel filtration elution peak was pooled, the protein concentration was measured by absorbance at 280 nm using a calculated extinction coefficient. 0.45 mL of the protein sample was put into each 0.5 mL MWCO 10K, 30K, 50K, and 100K Amicon filter. The samples were centrifuged at 8500 x g, 4°C until less than 50 µL remains on top of the filter. The total volume and protein concentration of each concentrated sample (concentrate) were measured. Any protein loss

was calculated for each filter by comparing protein amount on top of the filter with starting protein concentration.

Bottom portion of each filter (filtrate) was placed into a new 0.5 mL 10K Amicon filter. They were centrifuged at 8500 x g, 4°C until less than 50 µL remains on top of the filter. This concentrated all the detergent that passed through the initial filter so that it can be visualized on a TLC plate. Both concentrate and filtrate samples were kept for further TLC assay.

Detergent standards were prepared for a standard curve in the gel filtration buffer, including points above and below the expected detergent concentration. For example, standards of 0.25%, 0.5%, 1.0%, 2.0% (w/v) were prepared for an expected concentration of 0.5-1.0% corresponding detergent. 5 µL of each of the detergent standards, concentrate and filtrate samples were spotted onto the baseline of a silica 60 TLC plate. Pencil can be used to draw a straight line at the baseline (3 cm from the bottom of the plate) for aligned sample spotting and there should be at least 1 cm between sample spots. All samples were allowed to dry for at least 30 minutes at room temperature. Dried TLC plate was then placed into a TLC chamber containing about 0.5 inch solvent (chloroform:methanol:ammonium hydroxide, 63:35:5, v/v/v). The solvent was allowed to run at least half way up the plate. The TLC plate was removed from the chamber and the solvent on the plate was allowed to evaporate in open air for approximately 5 minutes. Dried TLC plate was placed in an iodine chamber (I_2) and allowed to stain for at least 5 minutes. The TLC was removed from the chamber and imaged immediately. The detergent amount was quantified by densitometry of the detergent spots using

ImageQuant, ImageJ, or Photoshop. Unknown detergent concentrations in concentrate and filtrate samples were estimated by using the detergent standard curve.

Optimal filter MWCO was determined by comparing the protein loss with the detergent loss for corresponding protein-detergent complex. The filter with the least detergent in the sample and the least protein loss should be the best choice.

2.3. Results

2.3.1. Archeal SPP orthologs

In higher eukaryotes, proteins might need post-translational modifications to be functional and active. This makes their recombinant bacterial expression challenging. In this thesis, archeal SPP orthologs were chosen to be expressed in *E.coli*. Archeal SPPs can be studied as a model system to higher eukaryotic SPPs and presenilin.

Three archeal SPPs were chosen: *Haloarcula marismortui* SPP (mSPP), *Halobacterium salinarum* SPP (hSPP) and *Methanoculleus marisnigri JR1* SPP (MCMSPP) (**Figure 2.4**). mSPP and hSPP were already cloned by Dr. Raquel Lieberman during her post-doctoral research. MCMSPP was cloned separately with Jennifer Johnson, a former lab member after a publication on MCMSPP was published (*71*). pet22b-(+) was chosen as expression plasmid which has N-terminal pelB sequence and C-terminal hexahistidine tag (**Figure 2.3**). N-terminal pelB is a leader sequence and it directs the recombinantly expressed protein to the membrane. Hexahistidine tag was used to purify the protein. During molecular cloning step, structure of MCMSPP was not known yet. Therefore, signal sequence prediction servers were used to predict a possible signal sequence on the N-terminal (*67*). Both constructs with and without predicted signal

sequence, MCMSPP and MCMSPP ΔN23 respectively, were cloned into the plasmid for expression and purification of MCMSPP. After crystal structure of an archeal SPP became available (PDB ID 4HYC) (*42*), it was revealed that predicted signal sequence was actually a part of TM1 of MCMSPP. Therefore, all efforts were focused on MCMSPP afterwards.

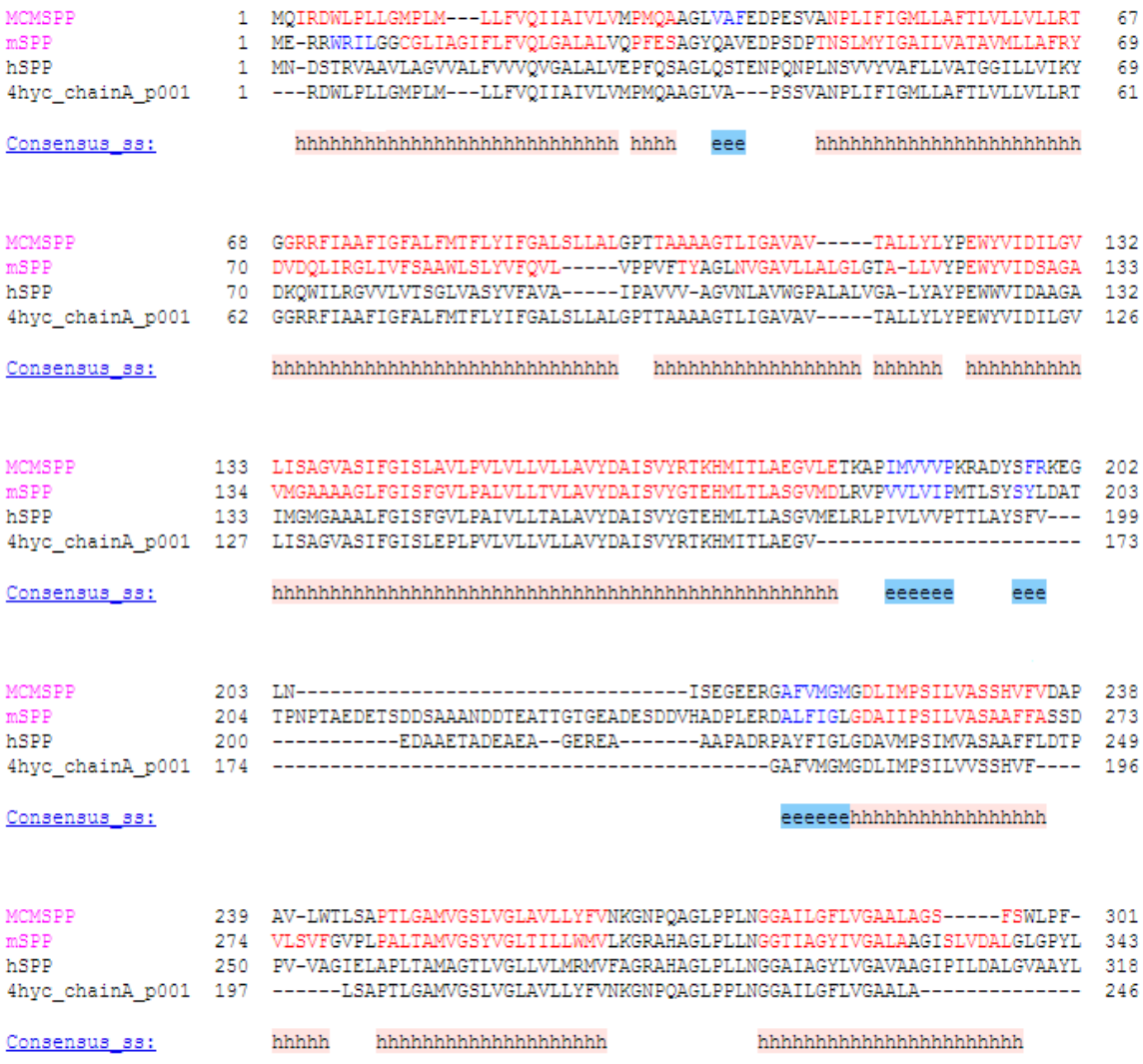


Figure 2.4. Multiple sequence alignment of MCMSPP, mSPP, hSPP and crystal structure sequence of MCMSPP (4hyc_chainA_p001, PDB ID: 4HYC). Alpha-helix and beta sheet residues are denoted as h and e respectively.

2.3.2. SPP expression, membrane isolation and purification

Three parameters were optimized for SPP expression: (i) induction temperature and duration, (ii) cell line, (iii) culture media. Induction at 18°C for 16-18 hours yielded more cell paste and higher protein expression than that of 37°C for 4-5 hours. *E.coli Rosetta 2* (DE3) gave better protein expression than *E.coli C43* (DE3). Although Luria-Broth (LB) and Terrific Broth (TB) gave similar cell growth curves (**Figure 2.5**), protein yield was higher with LB. This optimized protocol (Chapter 2.2.2) was used for all SPP orthologs. 2.5-3 g of cell paste was typically obtained from 1 L cell culture.

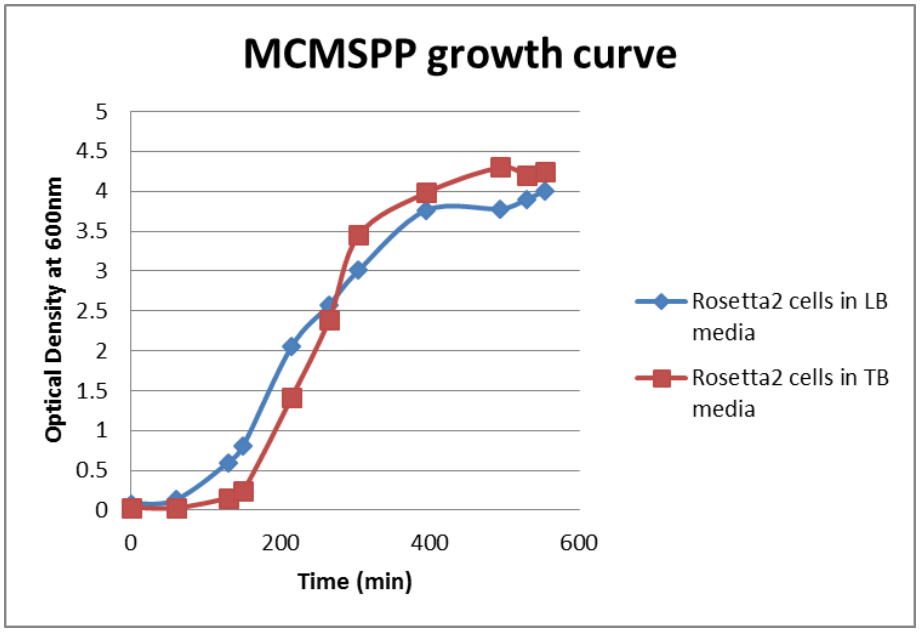


Figure 2.5. Cell growth curves for MCMSPP/*E.coli Rosetta2* (DE3) in different media.

Membranes were isolated from the cell paste as described in the methods section. A wash step was performed for the pelleted membrane to eliminate residual protease inhibitor. 0.9-1.2 g membrane was typically obtained from 7-8 g cell paste. Because membrane isolation itself takes a whole day, harvested membranes were always flash-frozen before resuspension and detergent solubilization.

SPP solubilization was performed with two different detergents: DDM and FC-12. They both solubilized SPP orthologs well enough for further purification steps. For purification, both DDM and FC-12 was tested. However, DDM was used for most of the purification protocols because FC-12 has not been successful for protein crystallization compared to DDM (*72*). For the last step of purification, SEC was used to exchange detergents. A variety of detergents (see Chapter 2.2.5) were tested to determine which detergent will be the most suitable for the final protein sample to be used in further structural and biochemical studies. First parameter to check was the shape of elution peak in SEC. mSPP with EE-tag (mSPP-EE1, Chapter 4) was chosen as a representative result (**Figure 2.6**). Gaussian-like peaks were preferred because it is an indicator of homogenous monodisperse SPP solution.

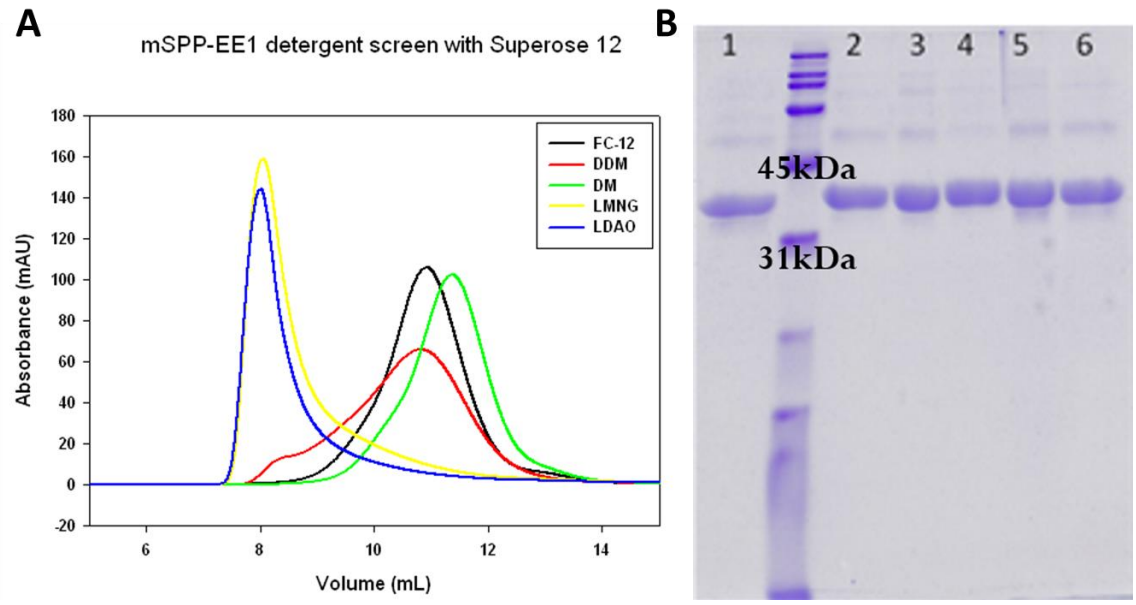
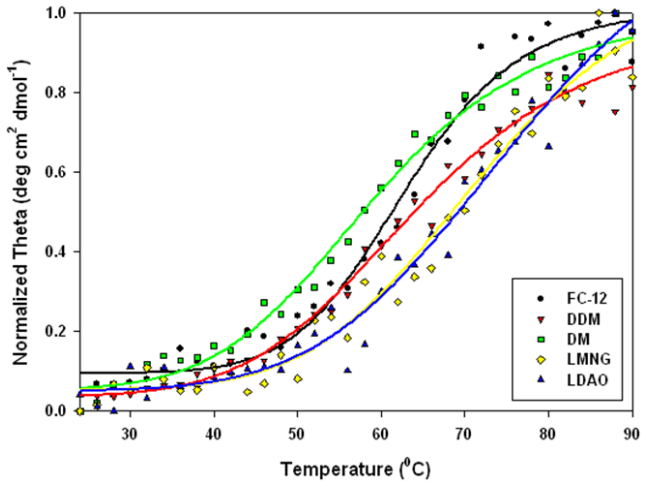


Figure 2.6. A representative example of detergent exchange by SEC. (**A**) SEC elution profile of mSPP-EE1 with 5 different detergents on Superose 12 column (**B**) SDS-PAGE of mSPP-EE1 in different detergents: 1) after Ni^{2+}-affinity purification with FC-12, after SEC with 2) FC-12, 3) DDM, 4) DM, 5) LMNG, 6) LDAO

2.3.3. Stability of SPP orthologs in different detergents

Stability analysis of SPP orthologs in different detergents were determined by CD thermal melt (**Figure 2.7**). Melting temperatures were calculated based on the melting curves. Some melting curves did not reach plateau due to the apparent high stability of membrane proteins and/or only partial unfolding in detergent micelles, extrapolation was performed to calculate their melting temperatures.

A



B

Detergent	T_m (°C)	Δ T_m
LDAO	72.8 ± 2.3	-
LMNG	69.6 ± 1.7	3.2
FC-12	62.5 ± 0.7	10.3
DDM	62.3 ± 1.2	10.5
DM	57.4 ± 0.8	15.4

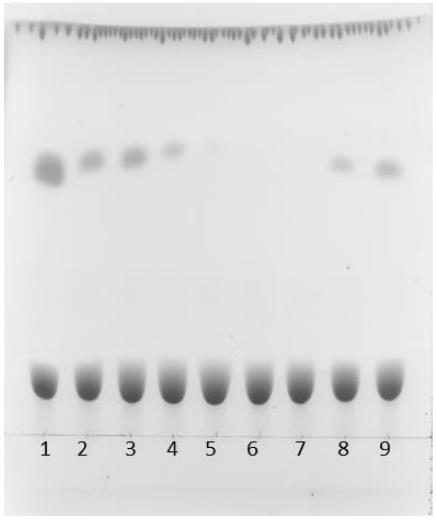
Figure 2.7. A representative example of CD thermal melt for stability determination. (**A**) CD thermal melt of mSPP-EE1 in different detergents. (**B**) Calculated melting temperature from CD thermal melt.

2.3.4. Detergent quantification and concentrator MWCO determination

Analysis of detergent quantity in the final sample was performed by TLC. Our aim was to eliminate as much empty detergent micelles as possible while retaining maximum amount of protein in the sample. Four different concentrator MWCO sizes were tried (10K, 30K, 50K, 100K, K: kDa). Both top (concentrate) and bottom (filtrate) of concentrator were analyzed by TLC for detergent quantification and UV-Vis

absorbance (at 280 nm) for protein quantification (**Figure 2.8**). 50K amicon filter was chosen for further use because it lost excess detergent and kept protein better than others.

A



B

	Protein percentage (%)	
Filter MWCO	Top	Bottom
10K	87.6	12.4
30K	86.9	13.1
50K	73.2	26.8
100K	44.3	55.7

Figure 2.8. A representative example of concentrator MWCO determination. (**A**) DDM detergent quantification of mSPP-EE2 samples in different concentrators by TLC analysis: 1) 3% DDM standard 2) 10K top 3) 30K top 4) 50K top 5) 100K top 6) 10K bottom 7) 30K bottom 8) 50K bottom 9) 100K bottom. (**B**) Protein percentages of mSPP-EE2 samples in different concentrators by absorbance analysis.

2.4. Discussion

In this chapter, three SPP orthologs were chosen to be expressed and purified for further biochemical and structural experiments. Several optimizations were performed to get better quality protein. Archeal SPP orthologs were chosen due to their ease of production in *E.coli* expression system. Cell growth and protein expression protocol was optimized to get highest yield of corresponding protein.

Membrane proteins need detergent for purification due to their hydrophobic nature. Different detergents were tested for initial solubilization and final protein sample. For solubilization, mostly DDM was tested first because it has been widely and successfully used for many other membrane proteins (*73*). Monodispersity and stability

are important parameters for downstream experiments (*74*). Detergent exchanges were performed at the last step of purification, SEC. Monodispersity was estimated from the shape of elution peak, gaussian peak shaped elution profile is an indication of monodispersity. For stability analysis, CD thermal melt was used. Melting temperature is a measure of protein stability which is related to crystallization propensity for structural studies (*75*). LMNG gave the highest stability to mSPP-EE1 according to CD melt data (**Figure 2.7**) but its SEC peak was in the void volume (**Figure 2.6**) meaning that its PDC size is higher than separation range of S12 (~300 kDa for globular proteins).

Detergent amount in the final sample is also important because excess detergent might interfere with crystallization (*76*). Different MWCO sizes of concentrator filters were tested. Concentrator with 50K MWCO was chosen for further experiments with DDM because it eliminated most of the excess detergent according to TLC analysis and it kept the protein better than 100K filter which lost more than half of the loaded protein (**Figure 2.8**).

Before TLC, we tried to use contact angle measurement (*77*) for quantification of detergent amount in the final sample (data not shown). However, we had many technical difficulties with the instrument which was located in another research lab (Behrens lab at Georgia Tech). In addition to those methods, there are several other techniques to determine detergent concentration such as refractive index, phenol-sulphuric acid assay for sugar-based detergents, molybdate assay for phosphate-containing detergents, radiolabeling, gas chromatography coupled with mass spectrometry, and nuclear magnetic resonance (*78, 79*). TLC was the most convenient method because it uses little

amount to sample, it works with most of the detergents and materials needed are not expensive.

Our method to determine stabilizing detergent was challenging because expression, purification, and CD melt should be performed for each SPP ortholog-detergent combination. There are other methods that can be used to make it more high-throughput such as fluorescent size exclusion chromatography (FSEC) with a GFP-fused protein (*80, 81*). In this method, protein even need not be purified but a fluorescence detector is required in line with the SEC column. Another method uses buried cysteines in your protein (if any) to perform high-throughput thermal unfolding experiment with N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM). CPM fluoresces upon reaction with thiol group on the cysteine which reveals with thermal unfolding (*82*).

This chapter summarizes our efforts to get higher yield, stability and quality SPP protein. We optimized our SPP expression and purification protocol accordingly, and this protocol has been used in multiple SPP related projects in the Lieberman lab.

CHAPTER 3: Development of Engineered Single Chain Antibody Fragments (scFvs) as Crystallization Chaperones for Membrane Proteins

3.1. Introduction

3.1.1. The use of crystallization chaperones for membrane protein crystallization

Macromolecular crystallization requires the association of weak, but specific, intermolecular interactions into a repeated lattice array, and is often the rate limiting step to structure determination by X-ray crystallography (*83*).

Although crystallization and structure determination of proteins is successful in many cases, numerous techniques have been implemented in which target proteins are modified to enhance crystallizability. The most well-known example is membrane proteins. Membrane proteins have a variety of essential functions in the cell, which are important from both biological and therapeutic standpoints. However, structure determination of membrane proteins is still challenging due to their hydrophobic nature. Techniques such as chaperone-based crystallization, tailoring the lipid and/or detergent environment, covalent modification, and mutagenesis studies have emerged to make crystallization of a target membrane protein tractable (*84-86*).

Chaperone-based crystallization uses co-crystallization chaperones that assist in the crystallization of a target membrane protein which present few surface polar residues to form crystal contacts (*86*). The addition of a chaperone protein can sequester aggregation-prone hydrophobic regions and can serve to immobilize flexible regions (*87, 88*). Structure determination of a G-protein coupled receptor (GPCR) using a fused T4

lysozyme in place of a flexible loop and a nanobody chaperone specific for the GPCR demonstrates the utility of this approach (*89*).

3.1.2. Antibody fragments as crystallization chaperones

Antibody fragments are the most commonly used chaperones for membrane protein crystallization (*85, 86, 90*). They include Fab fragments, single chain antibody fragments (scFvs) (**Figure 3.1**) and nanobodies (*89-92*). Although antibody fragment production is straightforward, it is still expensive and time-intensive (*86*). Additionally, produced antibodies are usually specific to a target protein; a more general approach would be beneficial. Our potentially generalizable approach uses engineered single chain antibody fragments (scFvs) specific to a short peptide epitope that can be inserted into any extramembranous loop of target membrane protein. Tight binding of engineered scFv variants to the membrane protein would provide a more hydrophilic crystal contacts leading to more stable crystal lattices, and better diffracting crystals.

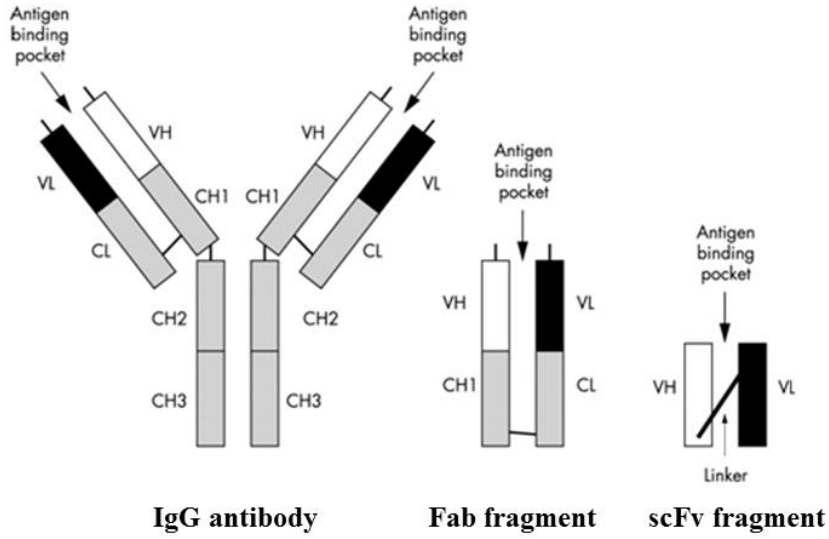


Figure 3.1. Fab and scFv antibody fragments derived from full Immunoglobulin G (IgG) antibody. Reprinted by permission from BMJ Publishing Group Ltd. (*93*).

3.1.3. Chapter overview and publications

To date, successful crystallization chaperones have been target-specific, limiting the generality of the chaperone method, as a new chaperone must be identified for each new target. Chaperone-assisted crystallization has the potential to become a general platform, however, if chaperones can be engineered for high affinity to a specific short peptide epitope that can be easily installed into any target protein of interest (**Figure 3.2**). We believe that an ideal chaperone toolbox would include multiple hypercrystallizable chaperones, each with specificity for a discrete epitope, versatile crystal contact interfaces and a predictable crystallization lattice to accommodate client proteins of various sizes. Exquisite control over crystallization may not be easily achievable, but for our set of engineered crystallization chaperones discussed in this chapter, we hypothesized that systematically exploring crystallization propensities may reveal governing principles and improved chaperones.

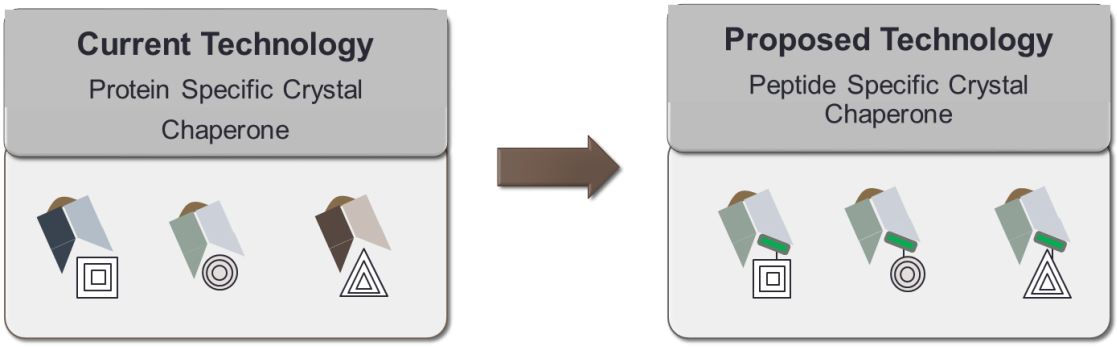


Figure 3.2. Target protein specific versus peptide epitope specific scFvs.

Here, we used a combination of computational analysis, protein design, crystallization, and structure determination to investigate the extent to which protein

crystallization may be controlled by tuning the energetics of intermolecular contacts within a highly conserved (sequence identity ≥ 85%) set of murine scFvs that are candidate crystallization chaperones. We previously engineered the anti-hexahistidine (His_6) scFv 3D5 (*94*, *95*) to enhance its biophysical properties and affinity for the hexapeptide EYMPME (EE-tag) (*95*), a process which inadvertently resulted in a new crystal lattice employing variable heavy chain (V_H) complementarity determining region (CDR) residues in a crystal contact. In this study, we systematically altered specific amino acid residues mediating these new crystal contacts in an attempt to restore the original 3D5 lattice, which does not use CDR residues in major crystal contacts. The experimentally prepared variants retain some crystal contact interfaces of the original parent 3D5 antibody, and in some cases even use the same residues for contacts; yet, lattices were altered. Taken together, our results underscore the challenges of directing a particular lattice in hypercrystallizable proteins such as this family of scFvs, but suggest that this plasticity could be an advantage for their use as crystallization chaperones.

In this work, protein engineering and design were mostly done by Maynard Lab at UT Austin, crystallization and structure determination were performed by Lieberman Lab and thermodynamic calculations were done by Truskett lab at UT Austin. This work was published in Proteins: Structure, Function and Bioinformatics in 2014 (*96*).

3.2. Methods

3.2.1. Molecular biology of scFv chaperones

Two initial scFv chaperones with enhanced biophysical properties were previously derived from 3D5 scFv (*94*): anti-His_6 3D5/His_683 and 3D5/EE_48, which has low nanomolar affinity for internal EE-tag (*95*). Two new scFv variants were

investigated in this thesis: anti-EE 3D5/EE_48.A and anti-EE 3D5/EE_48.K. Those new anti-EE variants had mutations which targeted crystal contacts in the lattice formation. 3D5/EE_48.A harbored the heavy chain amino acid changes S30T and S32A, and 3D5/EE_48.K harbored S30T and S32K. All scFv chaperones were designed by a collaborator, Maynard lab (University of Texas at Austin), they were cloned into the SfiI–SfiI site of pAK400 plasmid. Antibody residues are numbered according to the Kabat system (see **Table 3.1** for light chain and **Table 3.2** for heavy chain sequences). Design of 3D5/EE_48.A and 3D5/EE_48.K variants were performed by site-directed mutagenesis (SDM, Quickchange II, Agilent Technologies) with 3D5/EE_48 as a template. Primers for 3D5/EE_48.A scFv and 3D5/EE_48.K variants can be seen in **Table 3.3**. Sequences were verified by DNA sequencing (MWG Operon).

Table 3.1. Light chain crystal contact residues of 3D5, 3D5/His_683, 3D5/EE_48, 3D5/EE_48.A, and 3D5/EE_48.K. Residues are listed according to the Kabat numbering system. Crystal contact residues are shown in dark grey, and CDR regions highlight with light grey.

Kabat #	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
3D5	M	A	D	Y	K	D	I	L	M	T	Q	T	P	S	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S	S
3D5/His_683	M	A	D	Y	K	D	I	V	M	T	Q	T	P	S	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S	S
3D5/EE_48	M	A	D	Y	K	D	I	V	M	T	Q	T	P	S	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S	S
3D5/EE_48.A	M	A	D	Y	K	D	I	V	M	T	Q	T	P	S	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S	S
3D5/EE_48.K	M	A	D	Y	K	D	I	V	M	T	Q	T	P	S	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S	S

									V_L CDR1																						
Kabat #	27	27A	27B	27C	27D	27E	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
3D5	Q	S	I	V	H	S	N	G	N	T	Y	L	E	W	Y	L	Q	K	P	G	Q	S	P	K	L	L	I	Y	K	V	S
3D5/His_683	Q	S	I	V	H	S	N	G	N	T	Y	L	E	W	Y	L	Q	K	P	G	Q	S	P	K	L	L	I	Y	K	V	S
3D5/EE_48	Q	S	I	V	H	S	N	G	N	T	Y	L	E	W	Y	L	Q	K	P	G	Q	S	P	K	L	L	I	Y	K	V	S
3D5/EE_48.A	Q	S	I	V	H	S	N	G	N	T	Y	L	E	W	Y	L	Q	K	P	G	Q	S	P	K	L	L	I	Y	K	V	S
3D5/EE_48.K	Q	S	I	V	H	S	N	G	N	T	Y	L	E	W	Y	L	Q	K	P	G	Q	S	P	K	L	L	I	Y	K	V	S

	V_L CDR2																														
Kabat #	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
3D5	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T	D	F	T	L	K	I	S	R	V	E	A	E	D	L
3D5/His_683	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T	D	F	T	L	K	I	S	R	V	E	A	E	D	L
3D5/EE_48	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T	D	F	T	L	K	I	S	R	V	E	A	E	D	L
3D5/EE_48.A	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T	D	F	T	L	K	I	S	R	V	E	A	E	D	L
3D5/EE_48.K	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T	D	F	T	L	K	I	S	R	V	E	A	E	D	L

							V_L CDR3																								
Kabat #	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108						
3D5	G	V	Y	Y	C	F	Q	G	S	H	V	P	F	T	F	G	S	G	T	K	L	E	I	K	R						
3D5/His_683	G	I	Y	Y	C	F	Q	G	S	L	V	P	P	T	F	G	A	G	T	K	L	E	L	K	R						
3D5/EE_48	G	I	Y	Y	C	F	Q	G	S	L	V	P	P	T	F	G	A	G	T	K	L	E	L	K	R						
3D5/EE_48.A	G	I	Y	Y	C	F	Q	G	S	L	V	P	P	T	F	G	A	G	T	K	L	E	L	K	R						
3D5/EE_48.K	G	I	Y	Y	C	F	Q	G	S	L	V	P	P	T	F	G	A	G	T	K	L	E	L	K	R						

Table 3.2. Heavy chain crystal contact residues of 3D5, 3D5/His_683, 3D5/EE_48, 3D5/EE_48.A, and 3D5/EE_48.K. Residues are listed according to the Kabat numbering system. Crystal contact residues are shown in dark grey, and CDR regions highlight with light grey.

Kabat #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
3D5	Q	V	Q	L	Q	Q	S	G	P	E	D	V	K	P	G	A	S	V	K	I	S	C	K	A	S	G	Y	T	F	T	D
3D5/His_683	Q	V	Q	L	Q	Q	S	G	P	E	D	V	K	P	G	A	S	V	K	I	S	C	K	A	S	G	Y	T	F	T	D
3D5/EE_48	Q	V	Q	L	Q	Q	S	G	P	E	D	V	K	P	G	A	S	V	K	I	S	C	K	A	S	G	Y	S	L	S	T
3D5/EE_48.A	Q	V	Q	L	Q	Q	S	G	P	E	D	V	K	P	G	A	S	V	K	I	S	C	K	A	S	G	Y	S	L	T	T
3D5/EE_48.K	Q	V	Q	L	Q	Q	S	G	P	E	D	V	K	P	G	A	S	V	K	I	S	C	K	A	S	G	Y	S	L	T	T

	V_H CDR1																				V_H CDR2										
Kabat #	32	33	34	35	35A	35B	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	56A	57	58	59
3D5	Y	Y	M	—	—	N	W	V	K	Q	S	P	G	K	G	L	E	W	I	G	D	I	N	P	N	N	G	G	T	S	Y
3D5/His_683	Y	Y	M	—	—	N	W	V	K	Q	S	P	G	K	G	L	E	W	I	G	D	I	N	P	N	N	G	G	T	S	Y
3D5/EE_48	S	G	M	G	V	N	W	V	K	Q	S	P	G	K	G	L	E	W	L	A	H	I	Y	W	D	D	D	—	K	R	Y
3D5/EE_48.A	A	G	M	G	V	N	W	V	K	Q	S	P	G	K	G	L	E	W	L	A	H	I	Y	W	D	D	D	—	K	R	Y
3D5/EE_48.K	K	G	M	G	V	N	W	V	K	Q	S	P	G	K	G	L	E	W	L	A	H	I	Y	W	D	D	D	—	K	R	Y

	V_H CDR2																														
Kabat #	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	82A	82B	82C	83	84	85	86	87
3D5	N	Q	K	F	K	G	R	A	T	L	T	V	D	K	S	S	S	T	A	Y	M	E	L	R	S	L	T	S	E	D	S
3D5/His_683	N	Q	K	F	K	G	R	A	T	L	T	V	D	K	S	S	S	T	A	Y	M	E	L	R	S	L	T	S	E	D	S
3D5/EE_48	N	P	S	L	K	S	R	A	T	L	T	V	D	K	S	S	S	T	V	Y	L	E	L	R	S	L	T	S	E	D	S
3D5/EE_48.A	N	P	S	L	K	S	R	A	T	L	T	V	D	K	S	S	S	T	V	Y	L	E	L	R	S	L	T	S	E	D	S
3D5/EE_48.K	N	P	S	L	K	S	R	A	T	L	T	V	D	K	S	S	S	T	V	Y	L	E	L	R	S	L	T	S	E	D	S

								V_H CDR3																							
Kabat #	88	89	90	91	92	93	94	95	96	97	98	99	100	100A	100B	100C	100D	101	102	103	104	105	106	107	108	109	110	111	112
3D5	S	V	Y	Y	C	E	S	Q	S	—	—	—	—	—	—	—	G	A	Y	W	G	Q	G	T	T	V	T	V	S
3D5/His_683	S	V	Y	Y	C	A	A	S	S	P	Y	S	M	R	A	A	M	D	Y	W	G	Q	G	T	T	V	T	V	S
3D5/EE_48	S	V	Y	Y	C	A	R	R	G	G	S	S	H	Y	Y	A	M	D	Y	W	G	Q	G	T	T	V	T	V	S
3D5/EE_48.A	S	V	Y	Y	C	A	R	R	G	G	S	S	H	Y	Y	A	M	D	Y	W	G	Q	G	T	T	V	T	V	S
3D5/EE_48.K	S	V	Y	Y	C	A	R	R	G	G	S	S	H	Y	Y	A	M	D	Y	W	G	Q	G	T	T	V	T	V	S

Table 3.3. Primers for 3D5/EE_48.K and 3D5/EE_48.A.

Constructs	**Forward (F) and Reverse (R) Primers (5'-3')**
3D5/EE_48.A	F: ATGGGTGTGAACTGGGTTAAACAGAGTCCAGG R: CCTATAAGTGACTGGTGACGACCATACCCACACTTG
3D5/EE_48.K	F: ATGGGTGTGAACTGGGTTAAACAGAGTCCAGG R: CCTATAAGTGACTGGTGATTCCCATACCCACACTTG

3.2.2. Protein expression and purification

For all scFv chaperones, a standard heat-shock plasmid transformation into *E. coli C43 (DE3)* cells were performed and it was plated in an agar plate containing 34 µg/mL chloramphenicol (plasmid resistance). Luria-Bertani (LB, Fisher) and Terrific Broth (TB, Fisher) cultures used for the rest of the protocol contain the same antibiotic content. After overnight (16-18 hours) incubation of the agar plate at 37°C, a single colony was chosen for expression immediately or the plate was wrapped with a parafilm and stored in the 4°C refrigerator up to one week.

A single colony from transformation plate were added into 5 ml LB broth (starter culture) in the morning and incubated at 37°C for 6-8 hours with shaking at 225 RPM. The starter culture was inoculated into 500 mL TB in a 2 L baffled flask and grown overnight at 25°C with shaking at 225 RPM. Next morning, cells were pelleted at 4200 x g for 10 minutes at 4 °C, then the cell pellet was resuspended in 500 mL fresh TB in 2 L flask and incubated for 1 hour at 25 °C, 225 RPM. Protein expression was induced by the addition of 1 mM IPTG (Calbiochem) to each culture and shaking was continued at 225 RPM at 25°C for 4-5 hours. The cells were harvested by centrifugation at 4200 x g for 10 minutes. Cell pellet was put in a small plastic bag and flash-frozen in liquid nitrogen. They were stored at -80°C freezer until further use.

Cell pellet was resuspended in 10 mL resuspension buffer (0.1 M Tris pH 8.0, 0.75 M sucrose) per gram of cell pellet. 10 mg lysozyme (Sigma-Aldrich) was dissolved in 1ml resuspension buffer per gram of cell and it was added to resuspension. 7.5 mL of 1 mM EDTA were added slowly per gram of cell. The resuspension was stirred for 1 hour at 4 °C, then 1 mL of 0.5 M $MgCl_2$ was added per gram of cell and stirring continued for an additional 1 hour. After centrifuging for 20 minutes at 47,800 x g, supernatant was subjected to Ni^{2+}-affinity chromatography on an ÄKTA FPLC system (GE Healthcare) with wash buffer (20 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole) and elution buffer (20 mM Tris pH 8, 500 mM NaCl, 500 mM imidazole). Gradient elution (0-60% elution buffer) was used to elute corresponding scFv chaperone. scFv was further purified by preparative size exclusion chromatography using a Superdex 75 16/600 column (GE Healthcare) equilibrated with HBS buffer (50 mM HEPES pH 7.5, 150 mM NaCl) on an ÄKTA FPLC system. Both monomer and dimer peaks were collected and pooled together for further analysis and/or crystallization (see section 3.3.5).

3.2.3. Biophysical characterization

Protein purity and size were assessed by standard reducing sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Qualitative analysis of the oligomeric state distribution in solution at equilibrium (scFv monomer-to-dimer ratio) was determined by calculating the area under each elution peak from size exclusion chromatography using Unicorn software (GE Healthcare).

Protein solubility was determined by quantifying the concentration of soluble protein after concentration of the protein to ~20 mg/mL, incubation for four days at 4 °C, and centrifugation to pellet insoluble material. Thermal stability was evaluated by

differential scanning fluorimetry (DSF) (*97*). Briefly, purified protein (20 µl of 200µM) or buffer blank were mixed with Sypro Orange (1 µl of a 1:1000 dilution; Molecular Probes), heated in a Real Time PCR machine (ViiaTM7; Applied Biosystems) at increments of 0.96°C/min from 25°C to 90°C and analyzed with ViiaTM7 software (Applied Biosystems) for the melting temperature (T_m), the midpoint of unfolding. Protein solubility and thermal stability experiments were performed by the collaborator, Maynard lab (University of Texas at Austin).

3.2.4. Protein crystallization

Initial crystallization conditions for 3D5/His_683 were identified by sparse matrix screening (Hampton Research HR2-139 and HR2-138). For 3D5/EE_48.A and 3D5/EE_48.K, we initially used conditions that resulted in diffraction quality crystals of the parent 3D5/EE_48 scFv, but included several optimization steps to increase the diffraction quality of the daughter scFvs. Variables optimized included protein concentration, protein:drop ratio, temperature, buffer concentration, salt, and polyethylene glycol (PEG) molecular mass in the mother liquor. Ultimately, the best crystals of 3D5/His_683 (8 mg/ml in 10 mM HEPES pH 7.5, 150 mM NaCl, HBS) were grown at room temperature with 0.2 M KI, 0.001 M Guanidinium HCl, 18% PEG 8000 (Hampton Research). These crystals appeared in 2-3 days and grew to a maximal size of 120-150 µm within 1 week. Crystals of 3D5/EE_48.K (7.5 mg/ml in HBS) were grown at 4°C with 0.1 M Tris pH 8.5, 0.2 M $Li_2(SO_4)$, 3% 6-aminohexanoic acid, 24% PEG 8000; crystals appeared in 4-5 days and grew to a maximal size of 30-40 µm within 2 weeks. Crystals of 3D5/EE_48.A (7.5 mg/ml in HBS) were grown at 4°C with 0.1 M BisTris pH 6.5, 0.2 M $Mg(OAc)_2$, 21% PEG 8000, crystals appeared in 4-5 days and grew to a

maximal size of 20-30 μm within 3-4 weeks. All crystals were grown utilizing a 1:1 reservoir:protein ratio. David Heaner, Jr. (former undergraduate researcher in Lieberman Lab) helped with some of listed crystallization experiments.

3.2.5. Data collection, structure determination, and refinement

Crystals were mounted in nylon loops and flash-cooled with liquid nitrogen prior to synchrotron X-ray data collection. Glycerol was used as a cryo-protectant by supplementing the reservoir conditions with 30% (v/v) (3D5/His_683), 20% (v/v) (3D5/EE_48.A) and 25% (v/v) (3D5/EE_48.K). Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-ID at the Advanced Photon Source at Argonne National Laboratory (Darien, IL). Crystallographic data sets were processed with HKL2000 (*98*) and unmerged data were exported for structure determination by molecular replacement using Phaser (*99*). For 3D5/His_683, the search model was the 3D5 scFv (PDB ID 1KTR) whereas for 3D5/EE_48.A and 3D5/EE_48.K, the 3D5/EE_48 structure (PDB ID 3NN8) was used. The atomic models were fit to the respective electron density maps using Coot (*100*), and then iteratively refined with Phenix (*101*), taking advantage of parameterization specifically designed for low resolution structures (*102*). POLYGON (*103*) (Phenix) was used to evaluate model qualities of all structures and they were in the acceptable range (*103, 104*). Crystallographic statistics are presented in **Table 3.4**. Atomic coordinates and structure factors for 3D5/His_683, 3D5/EE_48.A and 3D5/EE_48.K have been deposited at the Protein Data Bank with access codes 4NKO, 4NKD, and 4NKM, respectively.

Table 3.4. Crystallographic data collection and refinement statistics for scFv chaperones.

	3D5/His_683 (4NKO)	3D5/EE_48.K (4NKM)	3D5/EE_48.A (4NKD)
Data collection			
Space group	$C222_1$	C2	F23
Cell dimensions			
a, b, c (Å)	60.59 104.92 284.41	103.26 92.26 142.71	275.67 275.67 275.67
α, β, γ (°)	90 90 90	90 110.86 90	90 90 90
Resolution (Å)[a]	42.2 - 3.49 (3.62 - 3.49)	46.1 - 3.71 (3.84 - 3.71)	38.6 - 3.3 (3.42 - 3.30)
R_{merge} [a]	0.09183 (0.4878)	0.1049 (0.2206)	0.1131 (0.2985)
R_{meas}	0.11	0.13	0.12
I/σI [a]	12.00 (2.97)	11.85 (4.30)	20.21 (6.97)
Completeness (%)[a]	93.10 (95.17)	84.81 (82.24)	99.96 (100.00)
Multiplicity [a]	3.3 (3.5)	2.0 (1.9)	7.6 (7.6)
Refinement			
Resolution (Å)	42.22 - 3.49	46.13 - 3.71	38.60-3.30
No. of reflections	11089	11363	26045
R_{work}/R_{free}	0.3084/0.3441	0.2636/0.2820	0.1950/0.2313
No. of molecules			
Protein residues	708	944	944
B-factor protein (A^2)	90.4	84.2	50.0
R.m.s deviations (rmsd)			
Bond lengths (Å)	0.004	0.003	0.003
Bond angles (°)	1.25	0.87	0.94
Ramachandran			
Favored (%)	92	95	97
Outliers (%)	1.6	0.75	1.0

[a] Values for the highest resolution shell are in parenthesis.

3.2.6. Computational analyses

PDBe Protein Interfaces, Surfaces and Assemblies (PDBePISA) (*105*) was used to rank and analyze crystal lattice contacts by surface area and energy, as well as catalog critical amino acids in crystal contacts based on their formation of H-bond or salt bridge interactions. After excluding peptide binding (ID4 and ID5 of 3D5) and the native heavy-light chain interface (ID1 of all scFv variants) within the scFv monomer, identified by the Complexation Significance score (CSS), the top three interfaces were considered major crystal contacts and used for further analysis.

To assess the overall thermodynamic stability of each scFv in its native or the 3D5 lattice, the three refined EE-peptide binding scFv monomer structures were fit into the P3221 space group using secondary structure matching in Coot (*106*). The overall free energy of each lattice was assessed by summing the energetic contributions of major crystal contacts, identified by PDBePISA as those with a surface area >50 $Å^2$ or $\Delta G >$ 0.84 kJ/mol and excluding biological interfaces. The free energy contributions of each contact are calculated as $\Delta G_{contact} = \Delta G_{interface} + \Delta G_{H\text{-}bonds} + \Delta G_{salt\ bridges}$, with $\Delta G_{interface}$ = the solvation free energy gain upon formation of the interface, calculated as difference in total solvation energies of isolated and interfacing structures, and the contributions of each hydrogen bond and salt bridge assessed at 2.09 and 1.25 kJ/mol respectively. The total free energy of lattice formation per scFv was obtained by dividing the energies by the number of monomers in each asymmetric unit. Thermodynamic stability calculations were performed by a collaborator, Truskett lab (University of Texas at Austin).

3.3. Results

3.3.1. Analysis of first-generation scFv chaperones (3D5/His_68, 3D5/EE_48)

The anti-His_6 3D5 scFv framework was selected as the initial basis for our crystallization chaperone toolbox since it is one of the few scFvs that does not use CDRs in major crystal contacts, instead directing these towards a large solvent cavity that could accommodate a membrane protein (*94*, *95*). Anti-His_6 3D5/His_683 varies from the parent 3D5 by 13 amino acid residues in the V_H CDR3 loop (**Table 3.2**) and it was shown that its biophysical characteristics are better than 3D5 (*95*). On the other hand, 3D5/EE_48 was engineered by further manipulating V_H CDRs to convert its specificity from anti-His_6 to anti-EE (**Figure 3.3**). 3D5/EE_48 was previously crystallized under identical conditions as 3D5, but used the cubic space group F23 as opposed to the $P3_221$ space group used by 3D5 (*95*). Lattice analysis reveals that 3D5 (**Figure 3.4a**) and 3D5/EE_48 (**Figure 3.4b**) both present large solvent channels, but are in distinct arrangements.

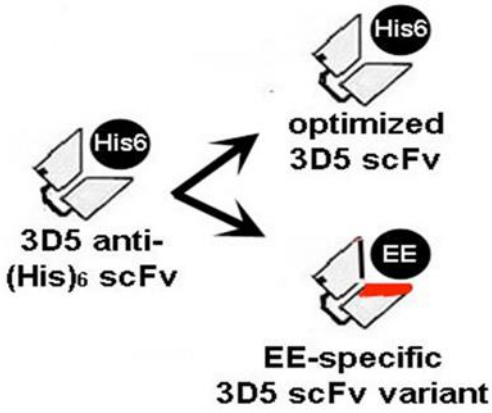


Figure 3.3. scFv engineering to improve its biophysical properties and change its affinity.

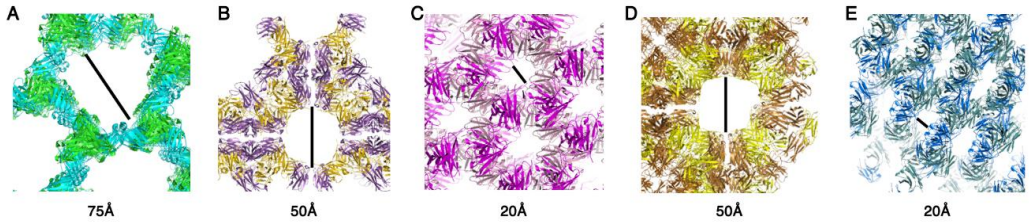


Figure 3.4. Crystal lattices of scFv variants described in this chapter. (A) 3D5, (B) 3D5/EE_48, (C) 3D5/His_683, (D) 3D5/EE_48.A, (E) 3D5/EE_48.K. Lines indicate solvent channels, with diameters listed.

Highly soluble 3D5/His_683 variant readily grew crystals in a variety of conditions. The final optimized 3D5/His_683 crystals grew in conditions reminiscent of, but different from, 3D5. No crystals were obtained in the original 3D5 conditions. The best diffracting crystals grew with a higher protein concentration and growth temperature, in a cocktail containing no buffer and a very low concentration of the denaturant guanidinium hydrochloride. The 3D5/His_683 structure was solved in space group $C222_1$ (**Table 3.4**). In this lattice, a pair of three nearly identical scFv molecules in the correct asymmetric unit is related by a two-fold rotation, but the structure could not be refined in a hexagonal lattice ($P6_522$ or $P6_122$ with one molecule in the asymmetric unit). These crystals exhibit significant pseudomerohedral twinning as identified by Xtriage within Phenix (*107*). Dozens of crystals were tested for space group assignment and reduced twin fraction, but no differences were detected. Lattice analysis reveals that the 3D5/His_683 lattice does not harbor an analogous large channel like 3D5 or 3D5/EE_48 to accommodate a client protein (**Figure 3.4C**). Although 3D5/His_683 does exhibit excellent biophysical properties (*95*) (see section 3.3.5), due to the combination of the crystal properties and disadvantages of targeting a pH-sensitive, C-terminal His_6 tag for

chaperoning, 3D5/His_683 was not considered further for membrane protein complexation (Chapter 4).

3.3.2. Structure and crystal lattice analysis of first-generation scFvs

Though they crystallize in different space groups, structures of the scFv monomers 3D5, 3D5/EE_48, and 3D5/His_683 are nearly superimposable (overall root mean squared deviation is ~ 1 Å), indicating there are no lattice-induced deformations (**Figure 3.5A**). PDBePISA crystal contact analysis (**Table 3.5**) reveals one common interface among the three structures (**Figure 3.5B**). This contact relies on polar interactions between C-terminal Lys and Arg residues within variable light chain (V_L) and other nearby residues of the V_H domain, which sometimes also involves a Gly from the linker region, depending on the variant (**Table 3.5**, **Table 3.1**, and **Figure 3.5C**). In 3D5, this contact is ranked third (ID3) and involves a surface area of 305.6 $Å^2$. It is likewise ID3 in 3D5/His_683 and ID4 in 3D5/EE_48. Unlike 3D5, which utilizes predominantly V_H-V_L protein contacts (**Figure 3.6A**), those in 3D5/EE_48 involve residues in symmetry related V_H domains, whereas for 3D5/His_683 there is a V_H-V_H contact and a V_L-V_L contact (**Table 3.5**).

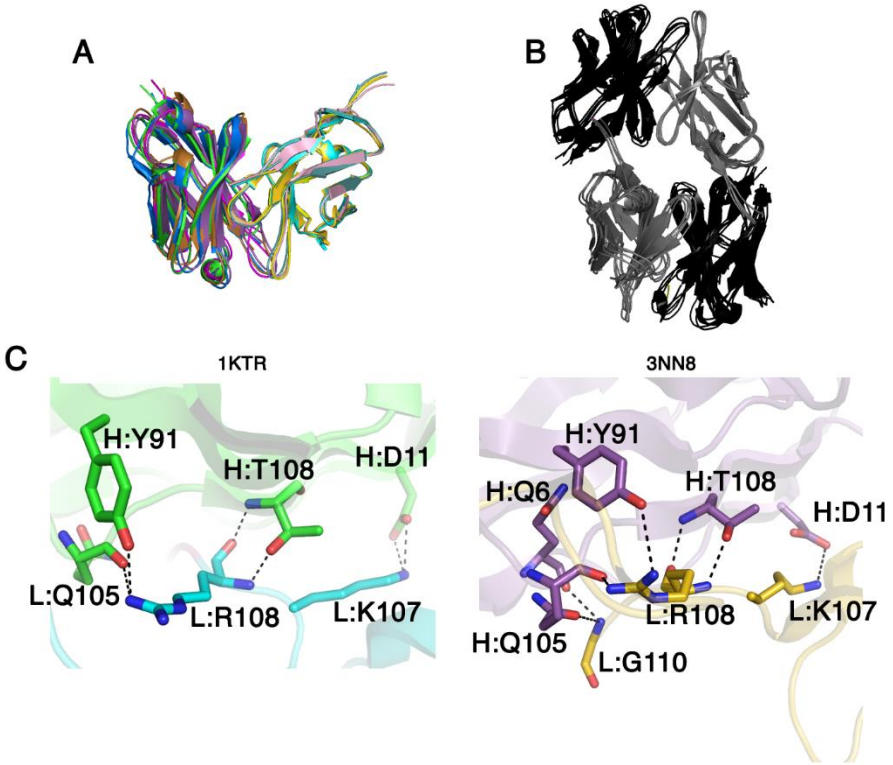


Figure 3.5. Common features among scFv variants. (A) Superposition of 3D5, 3D5/His_683, 3D5/EE_48, 3D5/EE_48.A, 3D5/EE_48.K monomers. (B) Common V_H-V_L crystal contact interface of aforementioned five structures (V_L in grey, V_H in black). (C) Close up view of contact in (B) for 3D5 (PDB code 1KTR) and 3D5/EE_48 (PDB code 3NN8). L: light chain, H: heavy chain.

Table 3.5. PDBePISA analysis of all scFv variants described in this study. In all variants, ID1 is the native V_L-V_H binding interface within the scFv monomer. Residues are listed according to Kabat numbering. See also Table 3.1 and Table 3.2.

scFv	PDBePISA Interface	Interface area (Å)	ΔG (kJ/mole)	V_H Contacts that formed H-bonds/salt bridges	V_L Contacts that formed H-bonds/salt bridges
3D5	ID2	467.1	11.3	T57, V71, D72, K73, S74	N30, K50, N53, R54
	ID3	305.6	-5.0	D11, Y91, Q105, T108	K107, R108
	ID6	160.0	10.0	n/a	Q42, R108
3D5/His_683	ID2	465.4	-2.1	D31, Y32, N52, S99	n/a
	ID3	353.0	-12.1	Q6, D11, Y91, Q105, G106, T108	L106, K107, R108, G109 (linker), G110 (linker)
	ID4	185.3	-2.5	n/a	S65, S67, D70
3D5/EE_48	ID2	465.4	3.3	K57, R58, S65, T68	n/a
	ID3	305.6	-16.3	P9, K19, S30, S32, D72, K73, S75, T77, Y79, S98	n/a
	ID4	393.0	-10.5	Q6, D11, Y91, Q105, T108	L106, K107, R108, G110 (linker)
3D5/EE_48.A	ID2	409.5	-15.5	P9, S21, T30, A32, T77, Y79, S98	n/a
	ID3	406.7	2.1	S65	n/a
	ID4	363.7	-12.6	Q6, D11, Q105, T108	K107, R108, G110 (linker)
3D5/EE_48.K[a]	ID2	483.7	-1.3	K23, S28, T30, D56, K57, Y59, K64, S65, S74, R82A	n/a
	ID3	356.3	-8.4	Q6, S7, D11, Q105, T108	K107, R108, G109 (linker), G110 (linker)
	ID4	226.0	-9.2	S98, Y100B	N28, N30

[a] Based on a low resolution (3.71Å) structural data

3.3.3. Rational design and analysis of second-generation scFvs (3D5/EE_48.A and 3D5/EE_48.K)

On the basis of the disparate crystal contacts found among 3D5 and the two first-generation scFvs (3D5/His_683 and 3D5/EE_48), residues were identified which, upon substitution with another residue, might restore the desirable 3D5 lattice (**Figure 3.4A**) via the original contacts (**Figure 3.6A**). Such residues were those involved in hydrogen bonding and/or salt bridge interactions spanning the intermolecular interfaces of 3D5, 3D5/His_683 and 3D5/EE_48, but were not involved in the common crystal contact described above (**Table 3.5**, **Table 3.6** and **Table 3.2**).

Five V_H residues, Ser 30, Ser 32, Lys 57, Arg 58, and Ser 65, were evaluated for their contributions to the respective crystal contact for each protein, and extent of accessible and buried surface areas. The top two residues, V_H Ser 30 and V_H Ser 32, found in 3D5/EE_48 ID3 and not predicted to interact with peptide (*95*), were then selected for mutation (**Table 3.6**). In 3D5/EE_48, Ser 30 occurs one residue before the start of V_H CDR1 (Kabat numbering) and participates in hydrogen bonding interactions with Thr 77 and Tyr 79 of a neighboring V_H domain via the Ser 30 side chain and main chain, respectively (**Figure 3.6B**). In this position in the structures of 3D5 and 3D5/His_683 there is a Thr, which is not involved in a crystal contact (**Table 3.5** and **Table 3.2**). V_H Asp 72 and V_H Lys 73 are involved in both 3D5 and 3D5/EE_48. Thus, by making the conservative S30T substitution, we expected to remove the 3D5/EE_48-specific Ser 30 interactions in favor of the desired V_H-V_L contact. In addition, in 3D5/EE_48, Ser 32 participates in hydrogen bonding interactions with the main chain carbonyl of Lys 19 from a neighboring V_H domain (**Figure 3.6B**). We reasoned that

replacement of Ser 32 with Ala would remove the side chain interaction, or substitution with Lys would disrupt the side chain interaction as well as increase entropy, making this contact less likely to form. Reversion to the Tyr 32 found in 3D5 and 3D5/His_683 did not seem prudent because while Tyr 32 does not form a crystal contact in the 3D5 structure, it does participate in a crystal contact in 3D5/His_683 and Tyr residues commonly mediate antibody-ligand interactions (**Table 3.5** and **Table 3.2**) (*108*). In sum, our second-generation scFv chaperone candidates are the double mutants 3D5/EE_48.A (S30T, S32A) and 3D5/EE_48.K (S30T, S32K).

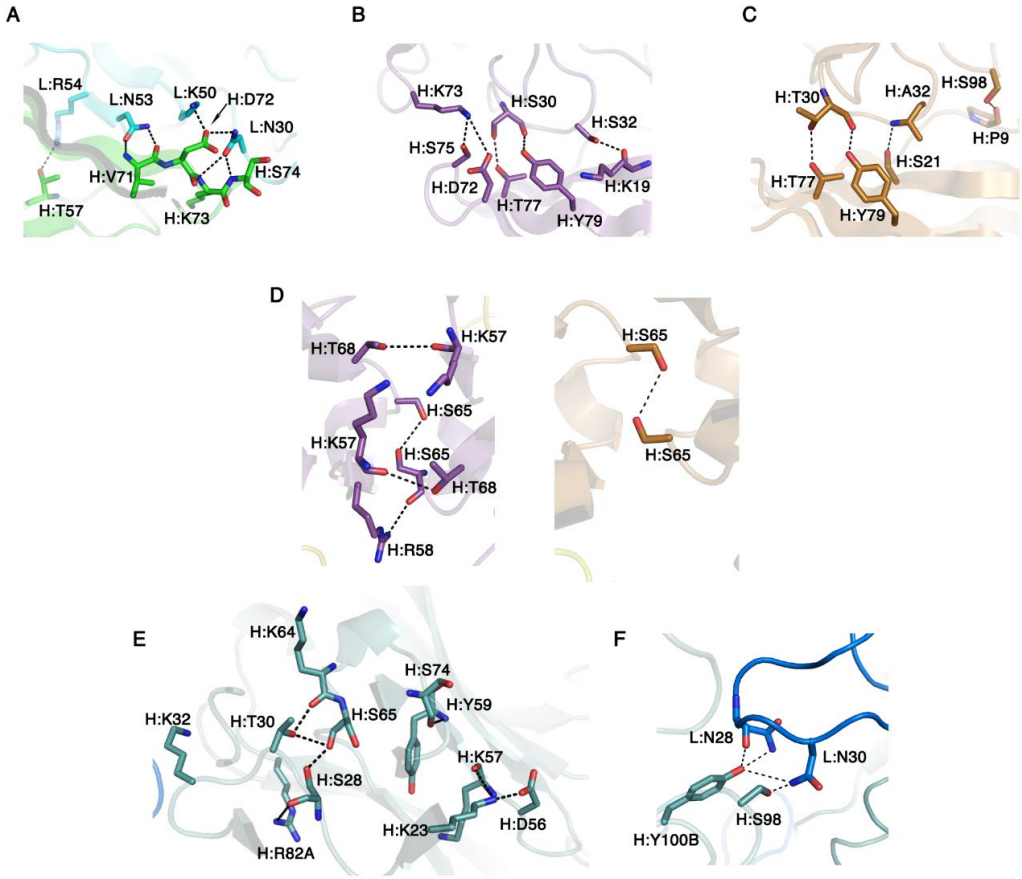


Figure 3.6. Outcome of rational design efforts. (A) Desired V_H-V_L lattice contact in 3D5 (ID2). (B) 3D5/EE_48 contact selected for mutagenesis (ID3) (C) Interface of 3D5/EE_48.A ID2 similar to 3D5/EE_48 ID3. (D) Comparison of 3D5/EE_48 ID2 (left) and 3D5/EE_48.A ID3 (right). (E) Interface ID2 formed in 3D5/EE_48.K (F) Interface ID4 formed in 3D5/EE_48.K. L: light chain, H: heavy chain. 3D5/EE_48.K interfaces are based on a low resolution (3.71 Å) structural data.

Table 3.6. Comparison of average solvent accessible surface areas of selected crystal contact residues.

Rank	Residue	Kabat #	PDBePISA Interface	ASA[a]	BSA[b]	% buried area[c]
1	Ser	30	ID3/7	46.1	44.7	96.9
	Ser	30	ID3/8	47.1	45.1	95.7
			Average	**46.6**	**44.9**	**96.3**
2	Ser	32	ID3/7	88.7	81.8	92.2
	Ser	32	ID3/8	82.0	76.0	92.6
			Average	**85.4**	**78.9**	**92.4**
3	Ser	65	ID2/5	91.8	83.5	91.0
	Ser	65	ID2/5	92.4	86.3	93.4
	Ser	65	ID2/6	94.1	86.5	92.0
	Ser	65	ID2/6	82.3	73.3	89.1
			Average	**90.1**	**82.4**	**91.4**
4	Lys	57	ID2/5	68.9	55.4	80.4
	Lys	57	ID2/5	79.3	59.9	75.4
	Lys	57	ID2/6	69.6	50.6	72.8
			Average	**72.6**	**55.3**	**76.2**
5	Arg	58	ID2/5	**151.6**	**68.0**	**44.8**

[a]Solvent-accessible surface area in $Å^2$, average values are in bold.

[b]Solvent-accessible surface area that is buried upon interface formation, in $Å^2$, average values are in bold.

[c]Buried area percentage (BSA/ASA*100) , average values are in bold.

3.3.4. Structure and crystal lattice analysis of second-generation scFvs

3D5/EE_48.A crystallizes in a cocktail nearly identical to that of 3D5 and 3D5/EE_48, but produces smaller crystals (20-30 μm in the largest dimension) than 3D5 (70 μm) (*94*) and 3D5/EE_48 (40-60 μm) (*95*). As with 3D5/EE_48 and 3D5/His_683, the scFvs are all readily superimposable (**Figure 3.5A**). Crystals of 3D5/EE_48.A belong to the same space group as 3D5/EE_48, namely, F23, with a similar 50Å solvent channel present in the lattice (**Figure 3.4D**). Interestingly, while V_H Ser 30 of 3D5/EE_48 and V_H Thr 30 of 3D5/EE_48.A are within essentially the same contacts in the two structures (**Figure 3.6B ,C**), V_H Ala 32 of 3D5/EE_48.A now torques the crystal contact to form a hydrogen bonding interaction with V_H Ser 21 via the main chain of V_H Ala 32 (**Figure**

3.6C), instead of interacting with V_H Lys 19 as seen for V_H Ser 32 in 3D5/EE_48 (**Figure 3.6B**). Notably, V_H Ser 21 is not a contact residue in 3D5/EE_48 (**Figure 3.6B**). Comparison of the second common contact (ID2 in 3D5/EE_48 and ID3 of 3D5/EE_48.A, **Figure 3.6D**) also reveals an apparent slippage between residues interacting across the contact. The only residue to form a contact at this interface in 3D5/EE_48.A is V_H Ser 65, which interacts with a symmetry-related V_H Ser 65, compared to four additional contacts in the same interface of 3D5/EE_48 (**Figure 3.6D**). Thus, the two amino acid substitutions were not sufficient to alter the crystal lattice but did change the lattice contacts.

By comparison, crystallization of 3D5/EE_48.K was successful under different conditions; in spite of considerable effort, no crystals were obtained in the condition used for 3D5. For 3D5/EE_48.K, the combination of higher protein concentration with a cocktail composed of monovalent rather than divalent salt, and higher pH, promoted crystal growth. The crystals belong to space group C2 (**Figure 3.4E**), and like those of 3D5/His_683 (**Figure 3.4C**), do not harbor a wide channel within the lattice. In the 3D5/EE_48.K structure, there are three main crystal contacts: (1) a hybrid interface (ID2, **Table 3.5**, **Figure 3.6E**) between the top two contacts seen in 3D5/EE_48 and ID2 in 3D5 (**Table 3.5**, **Figure 3.6A, B, D** left panel) (2) the common V_H-V_L interface (ID3) seen among all structures (**Table 3.5**, **Figure 3.5B** for overlay, **Figure 3.5C** close up of 3D5 and 3D5/EE_48), and (3) a new crystal interface (ID4) involving V_H CDR3 and V_L CDR1 (**Figure 3.6F**) proximal to ID2. Residues participating in 3D5/EE_48.K ID4 are located within V_H CDR3, the most important loop for peptide recognition, and V_L CDR1 (**Figure 3.6F**). One ID4 residue, V_L Asn 30, participates in the desirable common contact

in the 3D5 lattice (**Figure 3.6A**). In the hybrid contact of ID2, V_H Thr 30, one of the two residues altered in second generation variants due to its lack of involvement in contacts in the 3D5 or 3D5/His_683 lattices, participates in stabilizing interactions with the carbonyl oxygens of V_H Lys 64 and V_H Ser 65, both of which are residues within V_H CDR2 (**Figure 3.6E**). In turn, V_H Ser 65 is stabilized by V_H Ser 28, an interaction not observed in 3D5/EE_48 or 3D5/EE_48.A (**Figure 3.6D**). Residue 28 is an invariant serine within the EE-binding scFvs, a threonine in 3D5 and 3D5/His_683, and did not previously participate in crystal contacts (**Table 3.2**). Tyr 59 from V_H CDR2, which is unchanged in the structures under discussion and not previously in a crystal contact, forms a main-chain H-bonding interaction with the carbonyl group of V_H Ser 74 in a neighboring molecule (**Figure 3.6E**); V_H Ser 74 is a contact residue in 3D5 (**Figure 3.6A**) but not in 3D5/EE_48 or 3D5/EE_48.A (**Figure 3.6B, C**). Another commonality between the desired 3D5 contact and ID2 of 3D5/EE_48.K is the involvement of the residue at position 57 of V_H, which is a Thr in 3D5 and Lys in 3D5/EE_48.K (**Figure 3.6A, E**). Finally, a side-chain Lys-Asp salt bridge is present in both lattices (V_L K50-V_H D72 of 3D5 and V_H K23-V_H D56 of 3D5/EE_48.K), although specific residues are different (**Figure 3.6A, E**).

3.3.5. Biophysical characterization of second-generation scFvs

The minor changes in amino acid composition for 3D5/EE_48.A and 3D5/EE_48.K led to alterations in biophysical properties compared to first-generation chaperones 3D5/EE_48 and 3D5/His_683 (**Table 3.7**). Although all client proteins were purified in adequate yield for downstream experiments (**Figure 3.7**), the typical yield from cell growth of 3D5/EE_48.A was somewhat lower than 3D5/EE_48, and better

yield was obtained for 3D5/EE_48.K. Solubility was reduced for both new EE variants, but thermal stability was improved over 3D5/EE_48. Because solubility and yield are two very important factors for co-crystallization trials, 3D5/EE_48 and 3D5/EE_48.K are determined as the most promising anti-EE scFv crystallization chaperones (Chapter 4). Yield of 3D5/EE_48.A is very low to continue with downstream co-crystallization experiments, improvement of its yield through expression and purification optimization would not be a good effort since its low solubility (3mg/ml) would be another limiting factor for co-crystallization. As analyzed by size exclusion chromatography (SEC), the scFv variants 3D5/His_683 initially purify as a mixture of a monomer and dimer but to differing extents (**Figure 3.7A**). The anti-EE scFvs exhibit a higher monomer:dimer ratio compared to anti-His 3D5/His_683 and 3D5 (*95*) (**Table 3.7**), and the monomeric protein is stable once purified. Overall, the use of protein engineering has led to scFv variants with enhanced biophysical characteristics compared to 3D5.

Table 3.7. Biophysical characteristics of scFv variants.

	3D5/His_683[a]	**3D5/EE_48[a]**	**3D5/EE_48.A**	**3D5/EE_48.K**
Expression level (mg/l culture)	8.5	2.1	0.5	3.6
Solubility (mg/ml)	16.6	12.8	3	6.8
Melting temperature (°C)	53.6 ± 0.0	47.2 ± 0.3	51.75 ± 0.1	51.3±0.2
% Monomeric protein	62	79	72	75
Kd (nM), MBP-KEE (χ^2)	ND	166 (0.031)	100 (0.029)	147 (0.020)

[a] Pai, J. C., Culver, J. A., Drury, J. E., Motani, R. S., Lieberman, R. L., and Maynard, J. A. (2011) Conversion of scFv peptide-binding specificity for crystal chaperone development, *Protein Eng Des Sel 24*, 419-428.

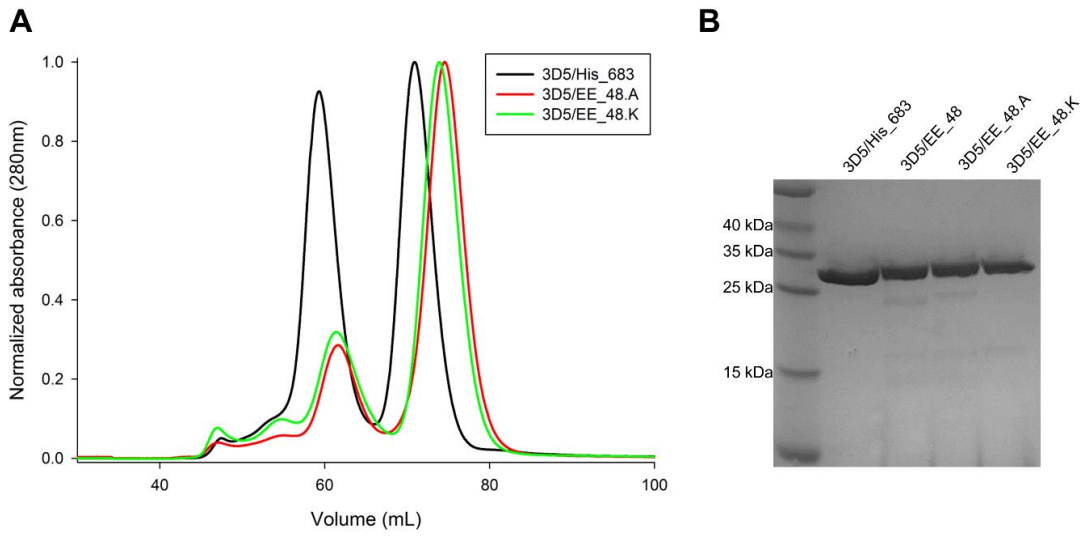


Figure 3.7. Biophysical characterization of second-generation scFv variants. (A) Oligomeric state (dimer:monomer) analysis of purified scFvs by size exclusion chromatography on Superdex 75 column. (B) Size and purity analysis of scFvs by SDS–PAGE.

3.3.6. Computational energetics analysis

To date, we have not crystallized our engineered anti-EE scFv variants in the 3D5 crystal lattice of $P3_221$ even though the residues mediating these interactions are conserved. To gain insight into why this may be the case, one of our collaborators, Truskett Lab at UT Austin, superimposed each engineered scFv onto the 3D5 lattice and computed the free energies for each lattice (**Table 3.8**). Comparable free energies are calculated among the anti-EE scFv variants in the hypothetical $P3_221$ lattice, similar to the value for 3D5. When compared to the experimentally-obtained lattices on a per-asymmetric unit basis, $P3_221$ is consistently less favorable, indicating that thermodynamics do contribute to lattice choice. However, when compared per scFv molecule, free energy values corresponding to the experimental lattices are either very

similar (3D5/EE_48 and 3D5/EE_48.K) or somewhat less favorable (3D5/EE_48.A) than $P3_221$ (**Table 3.8**).

Table 3.8. Free energy analysis of scFv variants in differing crystal lattices

scFv	Crystal Lattice[a]	ΔG^0 (kJ/mol asymmetric unit)	# scFv per asymmetric unit	ΔG^0(kJ/mol scFv)
3D5	**P 3_221**	**-18.8**	**1**	**-18.8**
3D5/EE_48	**F23**	**-70.3**	**4**	**-17.6**
	P 3_221	-15.9	1	-15.9
3D5/EE_48.A	**F23**	**-59.8**	**4**	**-15.0**
	P 3_221	-29.7	1	-29.7
3D5/EE_48.K[b]	**C121**	**-109.6**	**4**	**-27.4**
	P 3_221	-25.5	1	-25.5

[a] Experimentally observed lattice in bold.

[b] Based on a low resolution (3.71Å) structural data.

3.4. Discussion

We were motivated to understand crystallization behavior of highly similar scFvs as part of our effort to develop peptide-specific crystallization chaperones. Ideally, our anti-EE chaperones would template a porous three dimensional lattice using contacts remote from CDRs, in a manner suitable to immobilize an EE-tagged client protein for structure determination. We attempted to combine our knowledge of the crystallization propensity of 3D5, 3D5/His_683 and 3D5/EE_48 to redirect second-generation anti-EE scFvs to the 3D5 lattice, which uses largely non-CDR residues for contacts and whose lattice includes a wide ~75 Å diameter channel. Side chain entropy has been shown to be the principal negative correlate of crystallization, with charged residues the least likely to participate in crystal contacts (*109*). In the case of proteins that do not crystallize in their native state, mutation of key high-entropy side chains has been highly successful in identifying crystallizable variants (*110*). Thus, we used similar considerations of contact

residue contributions to buried and accessible surface area, as well as to surface entropy. The variants 3D5/EE_48.A and 3D5/EE_48.K retained potent EE binding affinity but crystallized in different, non-3D5 lattices. In the case of 3D5/EE_48.A, the lattice of 3D5/EE_48 was retained with minor modifications in crystal contacts. Our result for 3D5/EE_48.A is consistent with general observations from Surface Entropy Reduction (SER), namely that the substitution of Ser for Ala removes the polar side chain interaction and leads instead to a stereochemically constrained backbone interaction that likely contributes to hydrophobic stabilization in the water-depleted crystal contact environment (*110*). By contrast, our attempt to use SER underpinnings to kill a contact by introducing a lysine residue led to an unexpected result. The offending crystal contact from 3D5/EE_48 was altered, but a new hybrid contact was generated that incorporated some, but not all, of the desired parent contacts.

Our study demonstrates that the disruption of just one intermolecular contact can be sufficient to modify the lattice if more than one protein-protein interaction is available for crystallization. Such crystallization variability has been explored in classically studied proteins like lysozyme (*111-114*) and ribonuclease A (*115*), but few studies strategically attempt mutagenesis to alter the crystal lattice to a desired one starting from first principles. In one study, structures of *A. cellulolyticus* coh1 mutants were solved as part of an effort to generate a protein-based lattice capable of templating nanocomposite materials or as an immobilized catalyst (*116*). In this case, surface lysine residues were eliminated by replacing them with alanine or tryptophan. Although not a typical residue employed in SER (*110*), the latter mutant resulted in a crystal lattice with higher porosity and solvent content than the parent, with the Trp playing a key contact role (*116*). For

coh1, the presence of a surface-exposed Trp residue did not cause problems in terms of protein yield or solubility. In a second example, an alternative lattice was a serendipitous outcome of attempts to revert the oligomeric state of sweet potato β-amylase from monomer to the tetramer found in another ortholog (*117*). As in the case of our scFvs, residues comprising the 3D5 lattice contact were not mutated in our scFv variants to resurrect the desired crystal contacts. However somewhat different crystal contacts/interfaces were obtained, this shows that our scFv chaperones have versatile crystal contacts available for alternative lattice formations.

A major challenge in directing a particular crystal lattice is that beyond a general preference for amino acids with low entropic contribution (*110, 118*), physicochemical characteristics common to protein crystal contacts are not obvious (*112, 119*). In our system, one possible explanation for the observed lattice alterations is the length and flexibility of surface loops. Both 3D5/His_683 and 3D5/EE_48 have V_H CDR3 loops that are seven residue longer, while the 3D5/EE_48 V_H CDR1 is two residues longer, than in 3D5. These additional residues do not appear to directly influence crystal contacts, however, as just one of these additional residues (V_H Ser 98 in 3D5/EE_48 and V_H Ser 99 in 3D5/His_683) is involved in a crystal contact. Notably, while 3D5 exhibits the typical surface area for crystal packing of 100-500 Å (*112*) with a corresponding lattice free energy per asymmetric unit on the low end of published values (ΔG^0= ~-25-40 kJ/mol) (*83, 120*), those for our engineered anti-EE variants are higher in terms of contact surface areas, and are more energetically favorable per asymmetric unit. Viewed on a per-scFv basis, however, the free energies are roughly comparable; suggesting that lattice choice may be influenced by the crystallization pathway. Specifically, the additional residues in

our EE variants may direct crystal nucleation through repulsive or attractive events on the pathway to crystallization, such as by releasing key bound water molecules (*120*). The pathway may also be influenced by the composition of the mother liquor (*113, 115*) or the presence/absence of a ligand/binding partner.

Although controlling the crystallization lattice by rational design may not be achievable, our scFv platform holds considerable promise for crystallization chaperoning. As evidenced by this study, our hypercrystallizable scFv molecules have multiple well-ordered surface patches available for crystal contacts, including those utilized in 3D5 that would not interfere with epitope binding; indeed the presence of a bound ligand may independently direct lattice formation to that used by 3D5. In addition, our protein engineering efforts have led to a high affinity binding site, capable of complexation with a purified, detergent-solubilized, EE-tagged membrane protein (Chapter 4). Our chaperone discovery approach has advantages over current ligand specific library panning or hybridoma technologies in which numerous antibodies are evaluated for their ability to co-crystallize with a single client protein. With our approach, a single (or a few) chaperone(s) can be engineered for optimal crystallizability, expanded to nanobody and/or Fab platforms, and then used to co-crystallize a wide variety of client proteins simply by installing the hexapeptide epitope, the placement of which the researcher has complete control.

CHAPTER 4: Progress Towards Complexation and Co-crystallization of scFv Variants with EE-tagged Membrane Proteins

4.1. Introduction

4.1.1. Membrane protein crystallization challenge

Although membrane proteins are essential in many cellular functions and nearly one third of all encoded genes belongs to membrane proteins, their structural characterization is highly underrepresented. Less than 1% of unique protein structures in Protein Data Bank (PDB, www.rcsb.org) are membrane proteins (http://blanco.biomol.uci.edu/mpstruc/). The main reason behind this is their high hydrophobic content which makes crystallization challenging.

There are two main approaches to help membrane proteins to crystallize: (i) engineering membrane protein itself (*121*, *122*) or tailoring its lipid or detergent environment (*84*), (ii) use covalent or non-covalent crystallization chaperones (*85*, *86*). The most common technique that has been used is chaperone approach. Crystallization chaperones increase the hydrophilic surface available to form crystal contacts and therefore the likelihood of growing well-ordered and diffracting crystals. Covalent crystallization chaperones are usually inserted in one of the soluble loops of membrane proteins, one common example for this is lysozyme which is a highly-expressed, soluble and hypercrystallizable protein (*123*). Non-covalent crystallization chaperones form a tight complex with target membrane protein. Antibody fragments are the most common non-covalent chaperones because of their high affinity interactions (*85*, *86*). However, these approaches suffer from the need to identify a new chaperone specific for each new target protein to be crystallized. In our approach, we developed a generalized approach to

crystallize any membrane protein by altering a nonessential hydrophilic loop to harbor a short peptide epitope with high affinity for antibody crystallization chaperones. We chose EE-tag as the epitope which is a six amino acid long peptide EYMPME.

4.1.2. Advantages of EE-tag (EYMPME)

EE-tag was chosen as epitope due to its promising characteristics which would form a tight complexation (*124*). Tyrosine can form both hydrophobic and H-bond interactions which usually dominate protein-protein interactions. Proline can restrict conformational flexibility of the epitope, and two charged residues, glutamates, can form electrostatic interactions.

As described in Chapter 3, scFv antibody fragments were engineered for binding interaction to EE epitope while trying to retain the scaffold of parent 3D5 scFv known to use crystal contacts remote from the epitope binding site. We expect that our tightly bound scFv chaperone would immobilize EE-tagged membrane protein loop and provide a stable crystal lattice, leading to high quality diffracting crystals.

4.1.3. Chapter overview and publications

Membrane proteins are hard to crystallize and the developed techniques use target protein specific approaches to tackle this problem. In this chapter, a new method was tested as a more generalizable approach which used the high affinity interaction of anti-EE scFvs with EE-tagged membrane proteins. This chapter reports the successful tight complexations of several anti-EE scFv fragments with EE-tagged test membrane proteins. Both α-helical and β-barrel membrane proteins were tested for proof-of-concept

anti-EE scFv complexation. EE-tagged SPP and intimin were chosen as α-helical and β-barrel membrane proteins respectively.

The results from this chapter were published in Proteins: Structure, Function and Bioinformatics in 2014 (*96*). Another co-author paper which uses the same approach with anti-EE Fab antibody fragment was published in 2015 (*125*).

4.2. Methods

4.2.1. Site Directed Mutagenesis (SDM) for EYMPME (EE) tag substitution on extramembranous loops of membrane proteins

EE-tag epitopes were incorporated into proteins of interest via SDM (Quickchange II, Agilent Technologies) and verified by DNA sequencing (MWG Operon). Extramembranous loop between transmembrane helix (TM) 6 and 7 were chosen for EE-tag incorporation of SPP because it is the longest loop in all three chosen archeal SPP orthologs. EE-tags were installed on various locations of TM6-TM7 loops of MCMSPP, mSPP and hSPP except mSPP-EE3 whose EE-tag was installed on TM8-TM9 loop (**Figure 4.1**, **Table 4.1**). First EE-tagged SPP, mSPP-EE, was generated by Jeff Culver, a former Lieberman lab member, and it was also located in the TM6-TM7 loop (**Figure 4.1**, **Table 4.1**). Naming of EE-tagged MCMSPPs were shortened as MCM-EE(x) where (x) is the corresponding number of the EE-tagged variant.

E. coli intimin with the expression plasmid was generously provided by Dr. Susan K. Buchanan (NIH) (*126*). The EE-tag was incorporated into an extramembraneous loop in wild-type intimin (WT-intimin) between residues 314–321 via SDM (**Table 4.2**). More

details on EE-tag position optimization of intimin can be found in undergraduate thesis of David Heaner, a former undergraduate researcher in Lieberman lab.

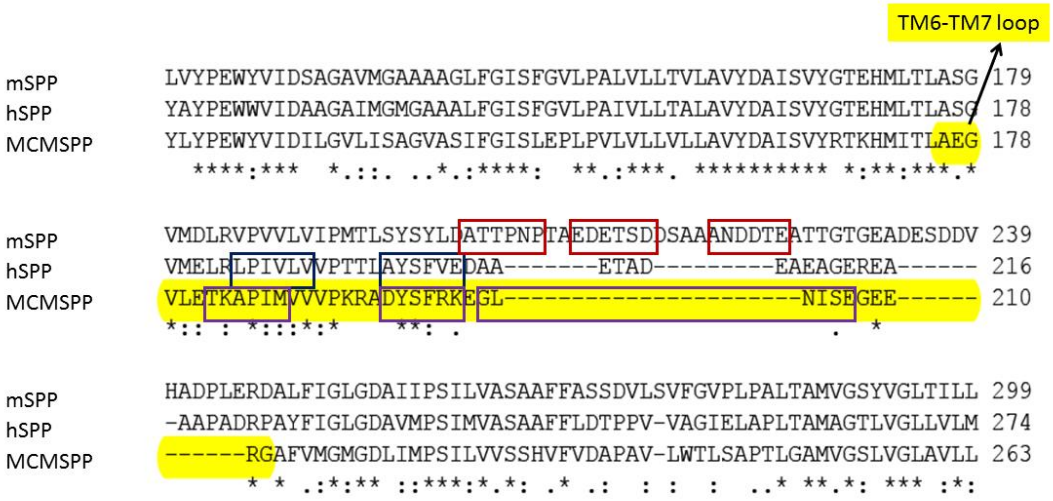


Figure 4.1. Multiple sequence alignment of TM6-TM7 loop locations of *Haloarcula marismortui* SPP (mSPP), *Halobacterium salinarum* (hSPP) and *Methanoculleus marisnigri JR1* SPP (MCMSPP). Loop was highlighted for MCMSPP. EE-tag substitution positions were labeled in rectangles for each (**Table 4.1**).

Table 4.1. Primers for EE-tag incorporation into mSPP, hSPP and MCMSPP.

Constructs with corresponding amino acid substitutions	**Forward (F) and Reverse (R) Primers (5'-3')**
mSPP-EE1 (TM6-TM7 loop) *(ATTPNP->EYMPME)*	F: CGCTGTCGTACTCTTATCTCGACGAGTATATGCCGATGGAAACAGCGGAAG ACGAGACGAGCG R: CGCTCGTCTCGTCTTCCGCTGTTTCCATCGGCATATACTCGTCGAGATAAGA GTACGACAGCG
mSPP-EE (TM6-TM7 loop) *(EDETSD->EYMPME)*	F: CGCCGAACCCAACAGCGGAATATATGACGAGCGATGACTCGGCG R: CGCCGAGTCATCGCTCGTCATATATTCCGCTGTTGGGTTCGGCG
mSPP-EE2 (TM6-TM7 loop) *(ANDDTE->EYMPME)*	F: CGAGCGATGACTCGGCGGCGGAGTACATGCCCATGGAGGCGACGACGGGC ACTGGCG R: CGCCAGTGCCCGTCGTCGCCTCCATGGGCATGTACTCCGCCGCCGAGTCATC GCTCG
mSPP-EE3 (TM8-TM9 loop) *(GRAHAG->EYMPME)*	F: GGATGGTGCTGAAGGAATACATGCCCATGGAACTTCCGCTTTTAAACGGTG GGACCATCGCCGGC R: GCCGGCGATGGTCCCACCGTTTAAAAGCGGAAGTTCCATGGGCATGTATTC CTTCAGCACCATCC
hSPP-EE1 (TM6-TM7 loop) *(LPIVLV -> EYMPME)*	F: GCCTCCGGGGTCATGGAGCTGCGGGAGTATATGCCGATGGAAGTTCCGACG ACGCTGGCGTACTCG R: CGAGTACGCCAGCGTCGTCGGAACTTCCATCGGCATATACTCCCGCAGCTC CATGACCCCGGAGGC
hSPP-EE2 (TM6-TM7 loop) *(AYSFVE -> EYMPME)*	F: TCGTTCCGACGACGCTGGAGTACATGCCCATGGAGGACGCCGCCGAGACG R: CGTCTCGGCGGCGTCCTCCATGGGCATGTACTCCAGCGTCGTCGGAACGA
MCM-EE1 (TM6-TM7 loop) *(TKAPIM->EYMPME)*	F: GCTGGCCGAAGGCGTCCTCGAGGAGTATATGCCCATGGAGGTCGTGGTTCC GAAGAGAGC R: GCTCTCTTCGGAACCACGACCTCCATGGGCATATACTCCTCGAGGACGCCTT CGGCCAGC
MCM-EE2 (TM6-TM7 loop) *(DYSFRK->EYMPME)*	F: GTCGTGGTTCCGAAGAGAGCGGAGTACATGCCCATGGAAGAGGGGCTCAA CATCAGTGAG R: CTCACTGATGTTGAGCCCCTCTTCCATGGGCATGTACTCCGCTCTCTTCGGA ACCACGAC
MCM-EE3 (TM6-TM7 loop) *(GLNISE->EYMPME)*	F: GACTACTCGTTCAGGAAAGAGGAGTACATGCCCATGGAGGGGGAGGAGCG CGGCGCGTTC R: GAACGCGCCGCGCTCCTCCCCCTCCATGGGCATGTACTCCTCTTTCCTGAAC GAGTAGTC

Table 4.2. Primers for EE-tag incorporation into WT-intimin.

	Forward (F) and Reverse (R) Primers (5'-3')
Intimin-EE	F: CGGCTACTTCCGTATGAGTGGTTGGCATGAATACATGCCCATGGAAGATTACG ATGAACGCCCGGCAAATGGCTTTGATATTCG R: CGAATATCAAAGCCATTTGCCGGGCGTTCATCGTAATCTTCCATGGGCATGTA TTCATGCCAACCACTCATACGGAAGTAGCCG

4.2.2. Purification of scFv fragments and EE-tagged membrane proteins

EE-tagged SPP was purified as described in Chapter 2.2 and scFv fragments (3D5/EE_48, 3D5/EE_48.A 3D5/EE_48.K) were purified as described in Chapter 3.2. EE-tagged intimin (intimin-EE) was expressed and purified as previously described for WT-intimin (*126*). More details on EE-tag position optimization and purification of intimin can be found in undergraduate thesis of David Heaner, a former undergraduate researcher in Lieberman lab.

4.2.3. Complexation of scFv fragments with EE-tagged membrane proteins

EE-tagged SPP was purified as described in Chapter 2.2 and scFv fragments (3D5/EE_48, 3D5/EE_48.A 3D5/EE_48.K) were purified as described in Chapter 3.2. EE-tagged intimin (intimin-EE) was expressed and purified as previously described for WT-intimin (*126*). More details on EE-tag position optimization and purification of intimin can be found in undergraduate thesis of David Heaner, a former undergraduate researcher in Lieberman lab.

The (SPP-EE)-scFv interaction was evaluated by size-exclusion-chromatography (SEC) using an ÅKTA FPLC instrument equipped with HiPrep 16/60 Sephacryl S-300 column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 250 mM NaCl, 0.05%

n-dodecyl-b-D-maltopyranoside (DDM, Anatrace), at 4°C. Wild-type or EE-tagged SPP was mixed in a 1:1 molar ratio with scFv and incubated for 1 hour on ice before injection. Elution fractions for each peak were analyzed with reducing SDS–PAGE.

The (intimin-EE)-scFv interaction was evaluated by SEC using an ÅKTA FPLC instrument equipped with a Superose 12 GL 10/300 column (GE Healthcare) equilibrated with 50 mM Tris pH 7.5, 200 mM NaCl, 0.01% Na Azide, 0.05% DDM, at 4°C. Wild-type or EE-tagged intimin was mixed in a 1:1 molar ratio with scFv and incubated for 3 hours on ice before injection. Elution fractions for each peak were analyzed with reducing SDS–PAGE.

4.2.4. Co-crystallization trials of scFv fragments with EE-tagged membrane proteins

After careful consideration of all SPP orthologs (Chapter 2), biophysical characteristics of scFv variants (Chapter 3), EE-tag positions and binding properties of SPP orthologs to scFv variants (Chapter 4), co-crystallization trials were focused on complexes of mSPP-EE and MCM-EE2 with 3D5/EE_48 or 3D5/EE_48.K (**Table 4.3**). Sparse matrices were prepared in 96-well format using the Art Robbins Crystal Gryphon instrument. Sparse matrix conditions were commercial screens purchased from Hampton Research, Emerald Biosystems, Rigaku, and Molecular Dimensions. If any crystal hit was observed in any sparse matrix trays, optimization trays were manually prepared in a 24-well format.

Two different methods were used to prepare protein sample to set up co-crystal trays: (i) concentrating complex peak from SEC with 10K MWCO amicon filters, (ii) directly mixing SPP and scFv in 1:1 molar ratio and incubating them on ice for 1 hour.

The latter method was mostly used for crystal trays that were set up. In total, 52 sparse matrix and manual optimization trays were set up for (SPP-EE)-scFv complexations and 65 trays were set up for (intimin-EE)-scFv complexations. More details on co-crystallization trials of EE-tagged intimin with scFv fragments can be found in undergraduate thesis of David Heaner, a former undergraduate researcher in Lieberman lab.

Table 4.3. SPP-scFv complexation and co-crystallization trials.

	Expressed/ Purified	Detergent Optimized	Complexation with 3D5/EE_48	Complexation with 3D5/EE_48.k	Co-crystallization
WT mSPP	Yes	Yes	Yes (-control)	Yes (-control)	-
mSPP-EE	Yes	No	Yes	Yes	Yes
mSPP-EE1	Yes	Yes	Yes	No	No
mSPP-EE2	Yes	Yes	Yes	No	No
mSPP-EE3	Yes	No	Yes	No	No
WT hSPP	Yes	Yes	No	No	-
WT-MCM	Yes	Yes	Yes (-control)	Yes (-control)	-
MCM-EE1	Yes	No	Yes	Yes	No
MCM-EE2	Yes	Yes	Yes	Yes	Yes
MCM-EE3	Yes	No	Yes	Yes	No

4.3. Results

4.3.1. Selection of EE-tagged membrane protein candidates based on SEC complexation performances

3D5/EE_48 and 3D5/EE_48.K were complexed with mSPP-EE and all three MCM-EE variants (**Table 4.3**). Their SEC profiles were compared and best SPP candidates were chosen. **Figure 4.2** shows the comparison of 3D5/EE_48.K complexations with WT-MCM/MCM-EE1/MCM-EE2/MCM-EE3. According to this comparison, MCM-EE2 was the best candidate for 3D5/EE_48.K complexation because intensity of its complex peak (around 60 ml) was higher and intensity of unbound 3D5/EE_48.K peak (around 90 ml) was lower than others. These peak intensities showed that more complex formed with MCM-EE2. Peak around 37 ml was void volume for this column and it might show the aggregated or higher molecular weight species but its amount was negligible compared to complex peak.

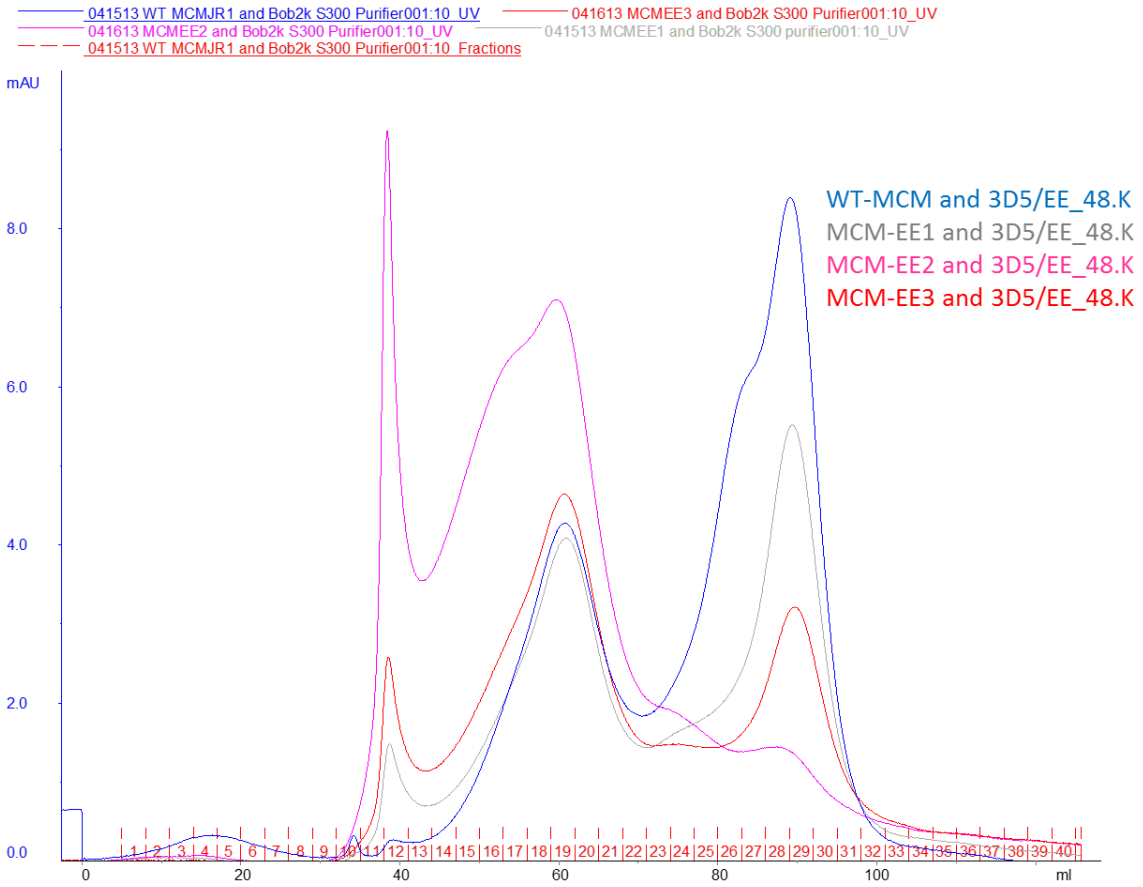


Figure 4.2. SEC complexation profile overlays of WT-MCM/MCM-EE1/MCM-EE2/MCM-EE3 with 3D5/EE_48.K on a HiPrep 16/60 Sephacryl S-300 column. y-axis is absorbance reading at 280 nm in mAU from AKTA FPLC instrument and x-axis is the elution volume in ml.

4.3.2. Complexation of EE-tagged SPP with anti-EE scFv fragments

Co-elution of 3D5/EE_48.K and MCM-EE2 was seen by SEC combined with SDS–PAGE analysis (**Figure 4.3**). There is a shift in the complexation peak (around 60 ml) towards the higher molecular weight (lower the elution volume, higher the molecular weight) for MCM-EE2 & 3D5/EE_48.K complex compared to negative control (WT-SPP & 3D5/EE_48.K). SDS-PAGE gel of corresponding fractions confirmed co-elution of 3D5/EE_48.K and MCM-EE2 in the complexation peak while WT-SPP and 3D5/EE_48.K eluted separately (**Figure 4.3**).

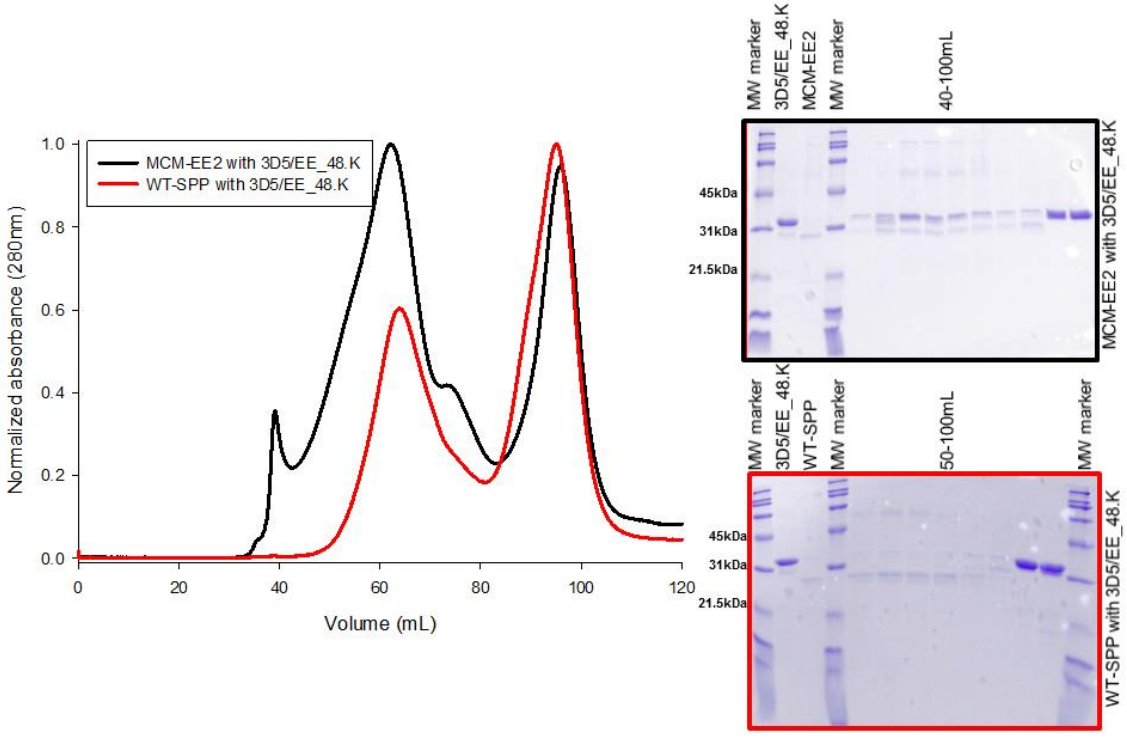


Figure 4.3. Overlay of MCM-EE2 & 3D5/EE_48.K (black) with WT-SPP & 3D5/EE_48.K (red) upon mixing and fractionation by SEC on a HiPrep 16/60 Sephacryl S-300. SDS–PAGE analysis of corresponding elution profiles are on the right; top: MCM-EE2 & 3D5/EE_48.K, bottom: WT-SPP & 3D5/EE_48.K.

4.3.3. Complexation of EE-tagged intimin with anti-EE scFv fragments

Co-elution of 3D5/EE_48.K and intimin-EE was seen by SEC combined with SDS–PAGE analysis (**Figure 4.4**). There is a shift in the complexation peak (around 12 ml) towards the higher molecular weight for intimin-EE & 3D5/EE_48.K complex compared to negative control (WT-intimin & 3D5/EE_48.K). This was also concomitant with a reduction in the height of the elution peak corresponding to unbound 3D5/EE_48.K (around 14 ml) when compared to the negative control. SDS-PAGE gel of corresponding fractions confirmed co-elution of 3D5/EE_48.K and intimin-EE in the complexation peak while WT-intimin and 3D5/EE_48.K eluted separately (**Figure 4.4**).

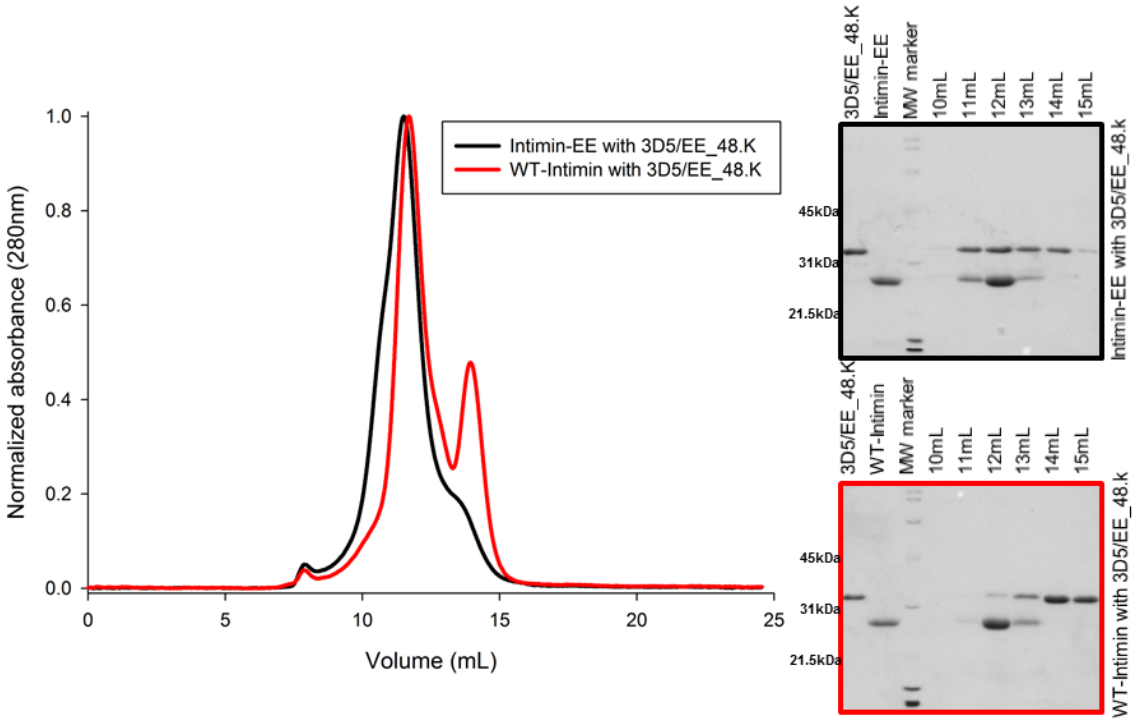


Figure 4.4. Overlay of intimin-EE & 3D5/EE_48.K (black) with WT-intimin & 3D5/EE_48.K (red) upon mixing and fractionation by SEC on a Superose 12 GL 10/300 column. SDS–PAGE analysis of corresponding elution profiles are on the right; top: intimin-EE & 3D5/EE_48.K, bottom: WT-intimin & 3D5/EE_48.K.

4.3.4. Co-crystallization trials

More than 50 co-crystal trays were set up for mSPP-EE/MCM-EE2 and 3D5/EE_48 and 3D5/EE_48.K complexes. Although there were some possible crystal hits, either they could not be reproduced with manual optimization trays or they were salt crystals confirmed by their diffraction pattern. After this, crystallization efforts were directed towards MCM-EE2 which formed diffracting reproducible crystals (Chapter 5).

Most of intimin-scFv or intimin-Fab co-crystal trays were prepared and screened by David Heaner, a former undergraduate researcher in Lieberman lab. David focused more on intimin-Fab co-crystallization due to the fact that purified scFvs were mixtures of monomers and dimers (Chapter 3) while Fab was mostly monomer and we could not

anticipate the effect of this in complexation and co-crystallization. Around 65 co-crystal trays were prepared for intimin-scFv and intimin-Fab complexes and more details can be seen in David Heaner's undergraduate thesis.

4.4. Discussion

Hybridoma technology and phage display are the most common techniques to generate antibody fragments specific to membrane protein of interest (*127, 128*). Although those approaches are widely used, they are still time-intensive, costly and a new antibody fragment needs to be generated for each new target protein. Another non-covalent chaperone that has been used in membrane co-crystallization is Designed Ankyrin Repeat Proteins (DARPins) (*129, 130*). DARPins are derived from ankyrin repeat which is 33 amino acid motif and shown to direct protein-protein interactions in the cell (*129*). Directed evolution strategies are used to generate DARPins specific for a target protein (*130*). All of those co-crystallization chaperone approaches suffer from targeting only single membrane protein of interest. Our generalizable approach use antibody fragments specific to a short peptide epitope which can be tightly complexed with any membrane protein with the epitope installed in one of its soluble loops.

The feasibility of our chaperone approach was demonstrated here by solution complexation of 3D5/EE_48.K with MCM-EE2 and intimin-EE, which are representatives of α-helical and β-barrel test membrane proteins respectively. To the best of our knowledge, this was the first demonstration of a rationally designed non-covalent complex between a candidate crystallization chaperone and a membrane protein harboring an installed internal peptide tag.

It was our hope that with our approach, a single (or a few) chaperone(s) can be engineered for optimal crystallizability, and then used to co-crystallize a wide variety of client proteins simply by installing the hexapeptide epitope via a single step. Because of its higher solubility and expression level, 3D5/EE_48 and 3D5/EE_48.K were assessed for complexation with proof-of-concept membrane proteins (Chapter 3). EE-tag was introduced into various positions of extramembraneous loops of candidate membrane proteins using just one round of site-directed-mutagenesis (SDM). EE-tag insertion did not change expression and purification characteristics of any test membrane proteins. The effect of EE-tag position was assessed by their complexation performances to anti-EE scFvs. It was shown that positions close to the middle region of loops are best for better complexation (mSPP-EE, MCM-EE2 and intimin-EE, section 4.2.1). Middle positions probably expose the EE-epitope towards the solution better for tight scFv binding interaction.

Although we set up many co-crystallization trays, we could not obtain any co-structures. To understand why the co-crystallization trials have not been successful in spite of the demonstrated favorable solution properties, we performed molecular dynamics in collaboration with Gumbart lab at Georgia Tech. WT-intimin and intimin-EE were first modelled and allowed to equilibrate over 50 ns. In comparison to WT-intimin, intimin-EE showed increased flexibility in residues 314–321, where the EE tag was substituted. Conformational flexibility is known to preclude crystallization, binding of antibody fragments to this floppy region might deter crystal contact formation and ultimately co-crystallization. More details on this result can be found in Jennifer Johnson's (former Lieberman lab member) PhD thesis and published work (*125*).

CHAPTER 5: Optimization and Protein Engineering for SPP to Improve its Crystallization and Diffraction: Lessons Learned from Contamination of Acriflavine resistance protein (AcrB) of *E.coli*

5.1. Introduction

Proteins with increased solubility and stability are less prone to interactions leading to undesirable off-pathways of aggregation or precipitation (*131, 132*). Biophysical properties can be improved through the use of libraries (*121*), modification of environment (*86*), optimization of target constructs to remove flexible regions (*133*), which reduces the probability of unproductive contacts, and truncation of loops that reduce conformational heterogeneity which also helps crystallization.

Reducing flexibility of extramembranous loops in membrane proteins is important for their crystallization because these soluble loops are expected to form the main crystal contacts if there are no available hydrophilic extracellular domains (*133, 134*). SPP does not have extracellular domains, but it has a very long loop between transmembrane (TM) helix 6 and 7 (Chapter 4). In the published crystal structure of SPP, this loop is not visible likely because Li, *et al.* used limited proteolysis approach to delete it in order to enhance crystallization (*42*) and thus could explain why the structure is in an inactive conformation.

5.2. Chapter overview and publications

In this chapter, several techniques were used in an attempt to enhance the crystallization of MCMSPP (Chapter 2). After getting the first crystal hit of MCMSPP, we optimized purification, crystallization, and post-crystallization steps to

reproduce/optimize crystals and improve its diffraction resolution. We also changed detergent environment of MCMSPP to lipidic cubic phases (LCP) and bicelles which would provide more native-like environment for crystallization.

Because initial crystals could not be reproduced without zinc salts, we hypothesized that zinc binding could be important for crystal contact interface. A zinc binding motif, HEXXH, was engineered into a soluble loop to make this possible interface even stronger which would improve crystallization properties.

We tried to shorten one of the longest extramembranous loops of MCMSPP by deleting amino acids step-by-step. By doing so, we aimed to observe the difference in crystallization properties of those SPP variants based on length of its loop. There are some conserved residues in the corresponding loop, we tried not to delete those residues to keep any important interactions. Activity assays were also done to test activity of those SPP variants to make sure they would crystallize in an active conformation.

Unfortunately, once the crystals were large and abundant enough to be visualized by silver staining after SDS-PAGE analysis, we learned that instead of crystallizing MCMSPP, we crystallized an *E. coli* membrane protein contaminant, acriflavine resistance protein (AcrB) which is a multidrug efflux pump. Because low amounts of AcrB nonspecifically binds to the Ni^{2+}-affinity column, we changed our constructs and purification protocol in hopes of eliminating this contaminant from our final protein sample. A TEV-cleavage site was added before C-terminal His_6-tag and a second affinity purification step was added to eliminate co-purified AcrB. A final strategy to remove AcrB was using *E. coli* strain without AcrB gene (*E. coli BL21 ΔAcrB*) for protein expression.

A book chapter was published in Methods in Molecular Biology volume titled Heterologous Expression of Membrane Proteins: Methods and Protocols, 2nd edition, in collaboration with former Lieberman lab member Dr. Jennifer Johnson (17). This manuscript covered some of the lessons learned from this study especially AcrB contamination issues.

5.3. Methods

5.3.1. MCM-EE2 crystallization trials

MCM-EE2 (Chapter 4) was first crystallized in MembFac sparse matrix (Hampton Research). Rod-shaped crystal hits were observed in H9 well with 7.3 mg/ml MCM-EE2 (0.1 M sodium cacodylate pH 6.5, 0.2 M zinc acetate, 9% PEG8k, 1:1 drop ratio).

In total, ~150 sparse matrix and manual crystallization trays were set up. In addition to protein engineering and environment optimization efforts, several other optimizations were done to improve crystallization as listed below:

- Concentration of protein (4 – 32 mg/ml)
- Different salt concentration in gel filtration (GF) buffer (50, 100, 150 mM NaCl)
- Concentration of each crystallization solution component
- Different polyethylene glycols (PEGs) for crystallization solution (PEG 400, 1000, 3000, 3350, 4000, 6000, 10000, 20000, PEG monomethyl ether 500, 2000)
- Different pHs for crystallization solution using suitable buffers (pH 5.5 - 8.5)
- Different salts (magnesium acetate, calcium acetate, ammonium acetate, lithium acetate, sodium acetate, ferric ammonium citrate, cobalt chloride, nickel sulfate,

copper acetate, cadmium chloride) and different zinc salts (zinc chloride, zinc nitrate) for crystallization solution

- Different temperatures for crystal tray incubation (4 °C, 20 °C, 16 °C and room temperature).
- Different detergents in GF buffer (Fos-choline-12, n-Dodecyl-β-D-Maltopyranoside, n-Decyl-β-D-Maltopyranoside, CYMAL-5, Lauryl Maltose Neopentyl Glycol, n-Dodecyl-N,N-Dimethylamine-N-Oxide, n-Nonyl-β-D-Glucopyranoside, n-Nonyl-β-D-Maltopyranoside, n-Octyl-β-D-Glucopyranoside, Anapoe-C12E8 from Anatrace). Minimum detergent amount which is 2x critical micellar concentration (CMC) for each detergent was used.
- Concentrator filter MWCO (10K, 30K, 50K, 100K) for final protein sample (Chapter 2)
- Different size-exclusion-chromatography (SEC) columns for GF: Superose 12 10/300 GL (S12) and HiPrep 16/60 Sephacryl S-300 HR (S300)
- Additive screens (additive screen and detergent screen from Hampton Research)
- Dehydration and glutaraldehyde-annealing of crystals right before flash-freezing
- Heavy metal and inhibitor (ZLL_2-ketone) soaking of the crystals

5.3.2. Crystallization trials with bicelles and Lipidic Cubic Phases (LCPs)

Bicelles were prepared as described in Ujwal, *et al.* (*135*), using 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (DMPC, Anatrace) and 3-[(3-Cholamidopropyl) dimethylammonio]-2-Hydroxy-1-Propanesulfonate (CHAPSO, Anatrace). A molar ratio of 2.8:1 (DMPC:CHAPSO) was chosen to prepare stock bicelle concentrations at 10%, 20% and 40%. Final protein sample for crystallization was prepared with 4-fold dilution;

one unit of bicelle stock was added to 3 units of purified concentrated protein sample on ice. All three stocks were used to do 4-fold dilution to obtain protein samples with final protein concentration of 18 mg/ml and final bicelle concentrations of 2.5%, 5% and 8%. Crystal trays were prepared on ice to maintain the fluidity of the bicelles. Crystal trays were incubated at 4, 20, or 37 °C. All crystal trays were imaged with the Rigaku Minstrel DT Benchtop protein crystal drop imager and recorded in CrystalTrak software.

Lipid cubic phase (LCP) crystallization trays were prepared in the Center for Structural Biology at The University of Alabama Birmingham. A total of 4 sparse matrix trays were prepared using the 1:1 direct mix method. In the final protein sample, protein:monoolein volume ratio was 60:40 and they were mixed by syringe lipid mixer (Hamilton Company) (*136*). Sparse matrix screens used were Cubic Phase I, Cubic Phase II, MB Class I and MB Class II (Molecular Dimensions). Hampton Research LCP sandwich plates were used and crystallization drops were prepared with the assistance of an Art Robbins Gryphon LCP instrument. The incubation temperature for the LCP trays was 20°C. All crystal trays were imaged with the Rigaku Minstrel DT Benchtop protein crystal drop imager and recorded in CrystalTrak software.

5.3.3. MCM-EE2 protein engineering via Site Directed Mutagenesis (SDM) to improve crystallization

5.3.3.1. Step-by-step TM6-TM7 loop shortening

The loop connecting TM6 and TM7 is the longest within SPP; in MCMSPP it is 37 residues long (**Figure 5.1**). This loop in MCMSPP was shortened step-by-step by SDM. At every step, 4 non-conserved amino acids in the loop were deleted from either MCM-EE2 or MCM-EE2-TEV construct (**Table 5.1**, Chapter 4, section 5.3.3.3). The

conserved predicted β-strand ("IMVVVP") at the start of this loop (**Figure 5.1**) as well as the EE-tag ("EYMPME"), were retrained (Chapter 4). The template used for all constructs was either the plasmid for ΔN23 or full-length MCM-EE2 in pet22b(+) plasmid (Chapter 2, Chapter 4, **Figure 5.2**). Construct sequences were confirmed by DNA sequencing (MWG Operon). Primers for all loop shortening steps are listed in **Table 5.1**.

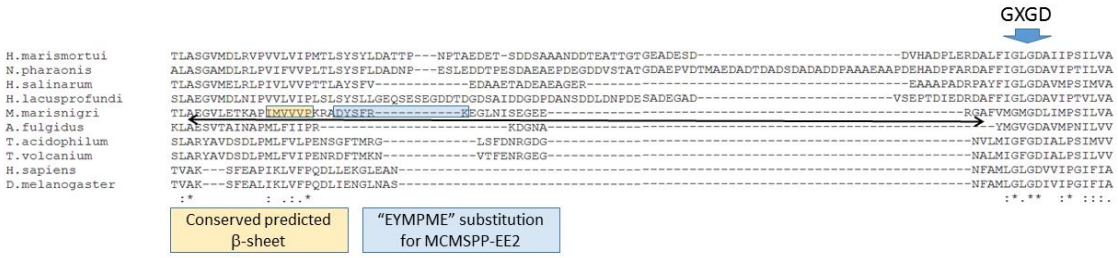


Figure 5.1. Multiple sequence alignment of TM6-TM7 loops of SPP from different organisms. The loop of MCMSPP was shown with an arrow. Conserved predicted β-sheet and EE-tag position for MCM-EE2 construct were highlighted in yellow and blue respectively. Conserved catalytic "GXGD" motif on TM7 was shown on top.

Table 5.1. Primers for stepwise MCM-EE2 loop shortening. E denotes for EE2, T denotes for TEV cleavage site, S denotes for loop shortening, the number after S denotes for number of rounds of 4 amino acid deletions (e.g. 1 denotes for 4 amino acids deletion, and 2 denotes for another 4 amino acids deletion after 1st round which is 8 amino acids in total)

Constructs	Forward (F) and Reverse (R) Primers (5'-3')
ES1/ETS1 (deletion of "EGEE")	F: GCCCATGGAAGAGGGGCTCAACATCAGTCGCGGCGCGTTCGTCATGGGTATGGGCG R: CGCCCATACCCATGACGAACGCGCCGCGACTGATGTTGAGCCCCTCTTCCATGGGC
ES2/ETS2 (deletion of "LNIS")	F: GCGGAGTACATGCCCATGGAAGAGGGGCGCGGCGCGTTCGTCATGGGTATGGGCG R: CGCCCATACCCATGACGAACGCGCCGCGCCCCTCTTCCATGGGCATGTACTCCGC
ES3/ETS3 (deletion of "ETKA")	F: CATGATCACGCTGGCCGAAGGCGTCCTCCCCATAATGGTCGTGGTTCCGAAGAGAGCG R: CGCTCTCTTCGGAACCACGACCATTATGGGGAGGACGCCTTCGGCCAGCGTGATCATG
ES4/ETS4 (deletion of "EGVL")	F: CGGACGAAACACATGATCACGCTGGCCCCCATAATGGTCGTGGTTCCGAAGAGAGCGG R: CCGCTCTCTTCGGAACCACGACCATTATGGGGGCCAGCGTGATCATGTGTTTCGTCCGG

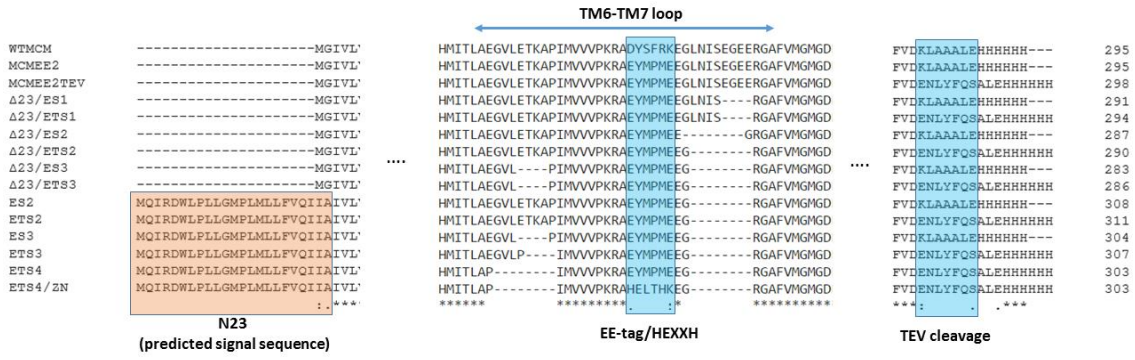


Figure 5.2. Multiple sequence alignment of all SPP variants that were studied for crystallization.

5.3.3.2. "HEXXH" motif on TM6-TM7 loop

The "HEXXH" peptide motif (where X is any amino acid) is known to form α-helix and accommodate zinc metal binding (*137*). From the initial crystallization efforts, it was shown that Zn^{2+} was essential for crystal formation. Therefore, we decided to substitute EE-tag region of ETS4 construct (**Figure 5.2**) with a "HEXXH" motif where X is any amino acid. "EYMPME" was mutated to "HELTHK", LT was chosen for XX region because it was common in natural α-helical HEXXH motifs, last amino acid of EE-tag, E, was mutated to K because WT MCMSPP originally had K in that position (**Figure 5.2**). Construct sequences were confirmed by DNA sequencing (MWG Operon). Primers can be seen in **Table 5.2**.

Table 5.2. Primers for "HELTHK" motif substitution of ETS4. EE-tag ("EYMPME") motif of FLETS4 was mutated into "HELTHK". Zn in the naming denotes for possible zinc binding property of "HELTHK" motif.

Construct	Forward (F) and Reverse (R) Primers (5'-3')
ETS4/Zn	F: CGTGGTTCCGAAGAGAGCGCACGAGCTGACCCACAAAGAGGGGCGCGGCGCG TTCGTC R: GACGAACGCGCCGCGCCCCTCTTTGTGGGTCAGCTCGTGCGCTCTCTTCGGAAC CACG

5.3.3.3. Tobacco etch virus (TEV) protease cleavage site construct and its additional purification step

TEV protease cleavage site ("ENLYFQS") was inserted before C-terminal hexahistidine (His_6) tag by SDM. Primers for TEV cleavage site insertion can be seen in **Table 5.3**. **Figure 5.2** shows which SPP variants were used for TEV cleavage site. Construct sequences were confirmed by DNA sequencing (MWG Operon).

Table 5.3. Primers for TEV cleavage site incorporation into SPP variants. Both insertion and substitution were made in two rounds of SDM for conversion of "KLAAALE" to "ENLYFQSALE".

Constructs	Forward (F) and Reverse (R) Primers (5'-3')
TEV round 1	F: GTTTTCGTGGCTTCCTTTCGTCGACGAGAATCTGTACGCACTCGAGCACCACC ACCACC R: GGTGGTGGTGGTGCTCGAGTGCGTACAGATTCTCGTCGACGAAAGGAAGCC ACGAAAAC
TEV round 2	F: CGTTTTCGTGGCTTCCTTTCGTCGACGAGAATCTGTACTTCCAGTCTGCACTC GAGCACCACCACCACCACCACTGAGATCCGGC R: GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCAGACTGGAAGTACA GATTCTCGTCGACGAAAGGAAGCCACGAAAACG

For MCMSPP variants containing a TEV cleavage site, removal of the hexahistidine tag was performed after the first Ni^{2+}-affinity purification step (see Chapter 2.2.4). The protein was buffer-exchanged using 50K MWCO Amicon Ultra centrifugal filter into gel filtration buffer to remove excess imidazole. TEV protease (~1.8 mg/ml in phosphate buffered saline, pH 7.5, 20% glycerol, stored at -80 °C freezer) was directly added to protein sample in a 1:1 TEV:MCMSPP variant mass ratio. The cleavage reaction was allowed to proceed at 4 °C for 16-20 hours and the sample was re-purified over a Ni^{2+}-affinity chromatography column, collecting the flowthrough and discarding protein that bound to the column (TEV protease and contaminants). Cleaved protein was concentrated and the gel filtration step was performed as in Chapter 2.2.4.

5.3.3.4. Modified T4 lysozyme fusion protein (mT4L) as a covalent chaperone for MCMSPP

Inspired by work of Kobilka in crystallizing GPCRs, a modified T4 lysozyme fusion (mT4L) was inserted into the TM6-TM7 loop of MCMSPP (*138*). A plasmid with mT4L sequence was ordered from Eurofins Genomics (**Table 5.4**). SacI and NheI

restriction sites were designed the N-term and C-term of ordered mT4L plasmid respectively. The same restriction sites were mutated via SDM into the ETS3 plasmid between the conserved β-strand and the start of TM7 (**Table 5.4**, **Figure 5.1**, **Figure 5.2**). Both mT4L and mutated ETS3 plasmids were double-digested with SacI and NheI, digested plasmids were cut/purified from the agarose gel and standard ligation protocol (NEB) was performed to obtain final SPP-mT4L plasmid (**Table 5.4**). Construct sequence was confirmed by DNA sequencing (MWG Operon).

Table 5.4. DNA/protein sequences and SDM primers for molecular biology of mT4L-SPP variant

mT4L sequence (*138*) ordered (SacI and NheI sites are underlined at the N- and C-terms respectively)	**DNA sequence:** GAGCTCATGAACATCTTCGAGATGCTGCGGATTGACGAAGGCGGTGGT AGTGGAGGCGATGAAGCCGAGAAACTGTTCAATCAGGATGTGGATGC GGCAGTACGCGGGATTCTTCGCAATGCGAAGCTGAAACCGGTGTATGA CTCGCTCGATGCTGTTCGTCGTGCAGCGTTGATCAACATGGTCTTTCAG ATGGGCGAAACTGGCGTTGCTGGGTTTACGAACAGCTTACGCATGCTG CAACAGAAACGCTGGGATGAAGCAGCCGTGAATCTGGCGAAATCCCG TTGGTACAACCAAACCCCAAATCGCGCCAAACGCGTCATTACCACCTT TCGTACGGGTACATGGGACGCTAGC **Protein sequence:** ELMNIFEMLRIDEGGGSGGDEAEKLFNQDVDAAVRGILRNAKLKPVYDSL DAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRW YNQTPNRAKRVITTFRTGTWDAS
Primers for SacI and NheI site SDMs into ETS3	**Forward (5'-3'):** CCATAATGGTCGTGGTTCCGGAGCTCGCGGAGTACATGCCCATGGAAG AGGGGGCTAGCGCGTTCGTCATGG **Reverse (5'-3'):** CCATGACGAACGCGCTAGCCCCCTCTTCCATGGGCATGTACTCCGCGA GCTCCGGAACCACGACCATTATGGGGACGCCTTCGGCC
SPP-mT4L protein sequence (mT4L sequence in TM6-TM7 loop is underlined)	MQIRDWLPLLGMPLMLLFVQIIAIVLVMPMQAAGLVAFEDPESVANPLIFI GMLLAFTLVLLVLLRTGGRRFIAAFIGFALFMTFLYIFGALSLLALGPTTAA AAGTLIGAVAVTALLYLYPEWYVIDILGVLISAGVASIFGISLAVLPVLVLL VLLAVYDAISVYRTKHMITLAEGVPIMVVVPELMNIFEMLRIDEGGGSGG DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGET GVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGT WDASAFVMGMGDLIMPSILVASSHVFVDAPAVLWTLSAPTLGAMVGSLV GLAVLLYFVNKGNPQAGLPPLNGGAILGFLVGAALAGSFSWLPFVDENLY FQSALEHHHHHH

5.3.4. Activity assays for MCM-EE2 shortened loop variants

Ren390FRET (R-E(EDANS)-IHPFHLVIHT-K(DABCYL)-R, Anaspec) was dissolved in DMSO (500 µM stock concentration). Freshly purified WT-MCM, ETS3 or ETS4 protein (0.5, 1, 2, 5 µM) in 10 mM sodium phosphate, pH 7.2, 150 mM NaCl (PBS), 0.05% (w/v) DDM and Ren390FRET (10 µM) in PBS with 0.05% (w/v) DDM, were mixed by gentle pipetting in a 96-well black-bottomed non-binding plate (Corning) and sealed with optical adhesive film (MicroAmp). The fluorescence values were monitored every 2 min for 1 hour at 37°C in a Synergy 2 plate reader (BioTek, filters λ_{ex}= 360 ± 40 nm, λ_{em}= 485 ± 20 nm). A blank was prepared without the enzyme and it was subtracted from the data. Initial velocities were calculated by the slope from the linear regression of data acquired in the first 20 minutes (AFU/min).

5.3.5. Protein expression with *E. coli BL21(DE3) ΔAcrB*

In an attempt to circumvent the AcrB contamination problem completely, we obtained the *E.coli* knock-out strain *BL21(DE3) ΔAcrB* from Dr. Edward Yu (Iowa State University) (*139*). AcrB is not essential in *E.coli* cell system. I used the same expression and purification protocol as explained in Chapter 2.

5.3.6. Two-dimensional (2D) crystallization efforts

In collaboration with Schmidt-Krey lab in School of Biology, 2D-crystallization trials of SPP for electron crystallography were performed. I purified different variants of SPP as in Chapter 2. Schmidt-Krey lab members, Maureen Metcalfe and Kasahun

Neselu, conducted 2D-cystallization trials and negative-staining visualization in the electron microscope.

Initially Ms. Metcalfe tested both purified mSPP and MCMSPP (Chapter 2). MCMSPP was identified as the more promising homolog due to its reproducible incorporation into lipid vesicles. Thereafter, electron crystallization trials focused on MCMSPP. Parameters tested to improve 2D crystals of MCMSPP include lipid-to-protein ratio, dialysis buffer content and pH, protein concentration, different fractions form SEC elution, purification protocol, and temperature. Additional experimental details can be found in Ms. Metcalfe's Master's thesis. Larger crystals with ordered areas and lower mosaicity are the main areas of improvement in this ongoing project to solve a high resolution structure of SPP in a native-like lipidic environment.

5.4. Results

5.4.1. MCM-EE2 crystallization

Our main goal was to co-crystallize EE-tagged SPP with anti-EE scFv (Chapter 4) but in addition to that, I set up some sparse matrix screens with left-over MCM-EE2 from co-crystallization trays. After 3-4 days, I had a crystal hit (**Figure 5.3A**). The hit condition was optimized several times by manual crystallization trays (section 5.3.1, **Figure 5.3B**). After getting crystals reproducibly, we decided to focus on MCM-EE2 crystallization in hopes of improving these crystals for structure determination. The same protein purification protocol was performed for the further crystal optimization trials (**Figure 5.4**).

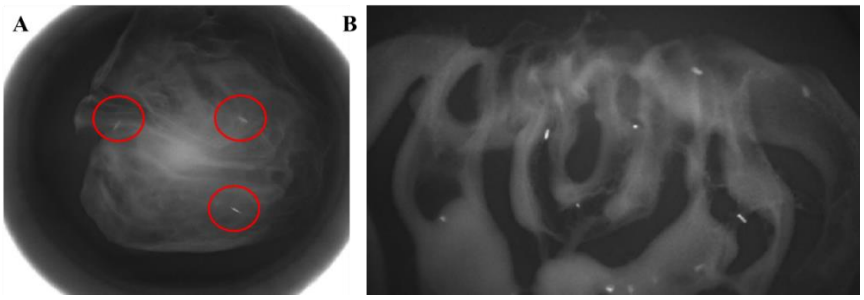


Figure 5.3. First crystal hit of MCM-EE2. (**A**) Crystal hits in H9 well of MembFac sparse matrix under UV imaging. Crystals are circled in red. H9 well solution is 0.1 M NaCacodylate pH 6.5, 0.2 M $Zn(OAc)_2$, 9% PEG 8000. 7.3 mg/ml MCM-EE2 was mixed with the solution in 1:1 ratio. (**B**) Optimization of sparse matrix crystal hit by manual crystal tray. Rod-shaped crystals are white under UV imaging. Well solution is 0.1 M NaCacodylate pH 6.5, 0.3 M $Zn(OAc)_2$, 5% PEG 8000. 7.1 mg/ml MCM-EE2 was mixed with the solution in 1 µL : 1 µL ratio.

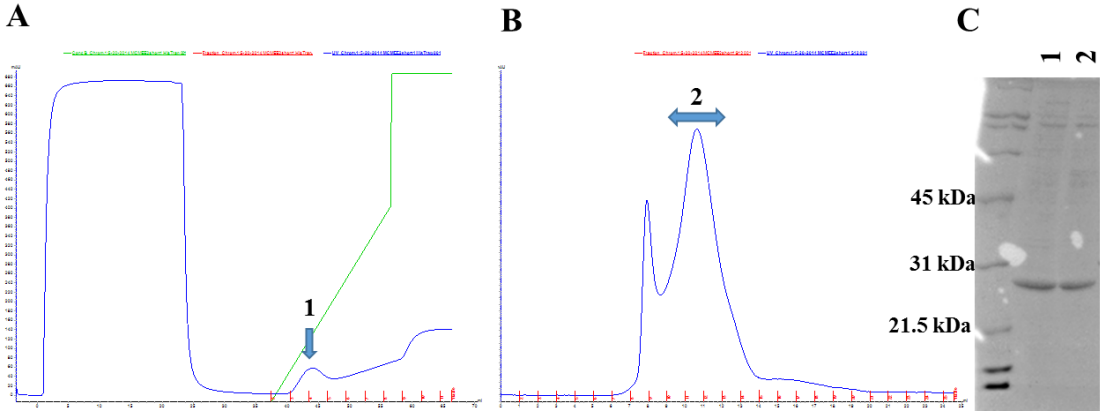


Figure 5.4. Representative purification flow of MCM-EE2 for crystallization optimization trials. (**A**) HisTrap purification with 0-60% gradient elution. Elution peak (1) are pooled and concentrated to ~250 µL for SEC. (**B**) SEC with S12 column, second peak (2) is pooled and concentrated for crystallization trials. (**C**) SDS-PAGE of HisTrap (1) and S12 (2) elution fractions. SPP band appears just below the 31 kDa band of molecular weight marker.

With optimization of the hit condition, the crystals got bigger and were subjected to X-ray diffraction data collection. These crystals diffracted to at best 12 – 7 Å (**Figure 5.5**) and were sensitive to radiation damage; the diffraction pattern decreased in intensity and diffraction limit, and the spots became very smeary during attempted data collection. Collected data were not of sufficient quality to determine the space group or unit cell

parameters. On the basis of these results, we decided to focus our optimization efforts on the crystallization buffer conditions and the construct itself, as explained in the upcoming sections.

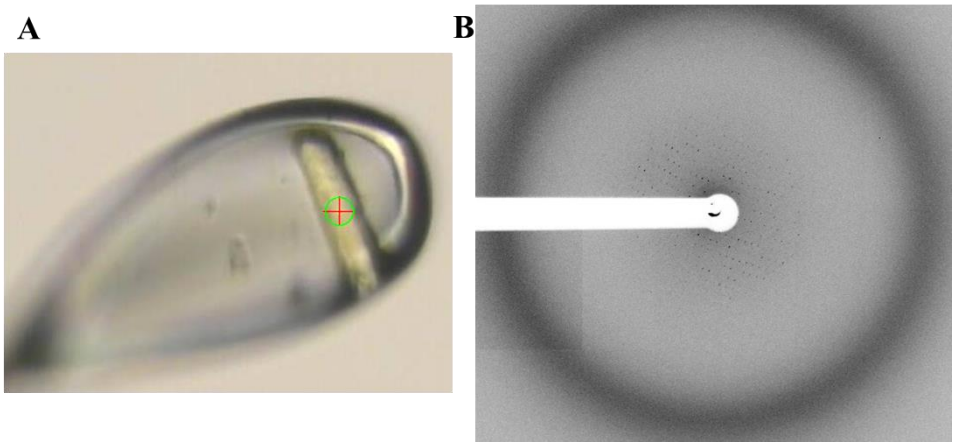


Figure 5.5. Representative MCM-EE2 crystal diffraction. (**A**) MCM-EE2 crystal flash-frozen in a nylon loop for X-ray diffraction. The crystal is grown in solution of 0.1 M NaCacodylate pH 6.5, 0.3 M $Zn(OAc)_2$, 5% PEG 8000 and cryo-protectant is the same solution supplemented with 25% glycerol. 10 µM beam size is represented as green circle on the crystal. (**B**) X-ray diffraction pattern of the crystal. Diffraction spots are observed up to 7 Å resolution.

5.4.2. Optimization of MCM-EE2 crystals

As summarized in Section 5.3.1, many parameters were tested in attempt to improve diffraction quality of MCM-EE2 or its variants. Many of optimization strategies were also employed with promising engineered MCM-EE2 variants (Section 5.3.3). **Figure 5.6** shows a representative result from filter MWCO and detergent additive screen optimization efforts. All crystals obtained throughout this optimization process were rod-shaped and improvement criteria were based on morphology (longer and thicker), which were expected to diffract to higher resolution.

Throughout the optimization process, zinc ions and a buffer of pH 6.5 were required to grow crystals; when I changed salt, no crystals grew. Crystals did not grow at all or were very small when the buffer used was lower or higher than pH 6.5. Additive screens and the post-crystallization optimizations (dehydration, annealing, and heavy metal/inhibitor soaking) did not have any substantial change in the crystallization. Crystal size and diffraction quality were improved with higher protein concentration in the crystal drop: the length of crystal became longer than 100 μm and they became thicker when the protein concentration was greater than or equal to 15 mg/ml.

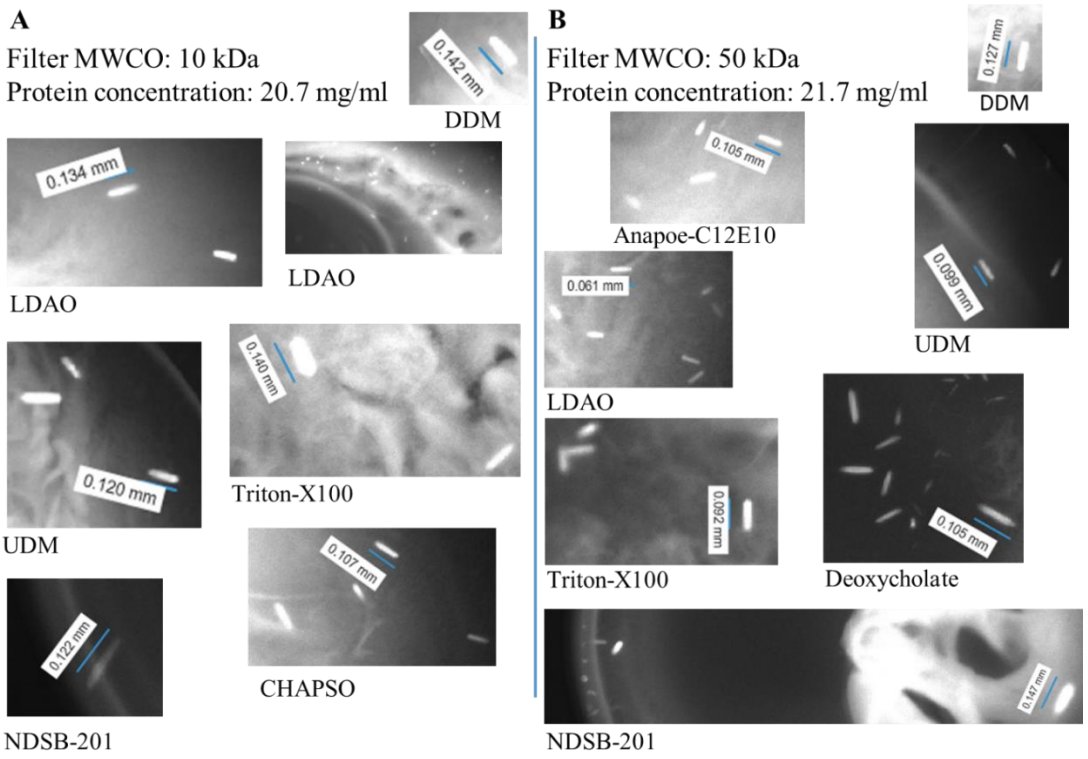


Figure 5.6. Representative crystal images from optimizations of filter size and detergent additive screen. MCM-EE2 protein is concentrated by (**A**) MWCO 10K and (**B**) 50K amicon filter for additive detergent screening. Various crystal hits in different additive detergents are shown. Crystallization condition is 0.1 M NaCacodylate pH 6.5, 0.3 M $Zn(OAc)_2$, 6% PEG 8000. Protein : additive detergent : crystallization solution volume ratio was 5:1:4.

Data collected on these larger crystals diffracted to ~4.5 Å resolution (**Figure 5.7**). Although radiation damage precluded collection of a complete dataset, we were able to index the initial frames obtained. Unfortunately, a search of the refined space group and unit cell dimensions matched with AcrB structures deposited in the Protein Data Bank (PDB). Though documentation as such is scant, AcrB is a relatively common contaminant in membrane protein crystallization trials. AcrB crystallizes even at picogram levels where it is not visible by SDS-PAGE (*140-142*). Based on this result, we inferred our protein crystals were AcrB and we turned our efforts to eliminate AcrB contamination (sections 5.4.3, 5.4.5, 5.4.6).

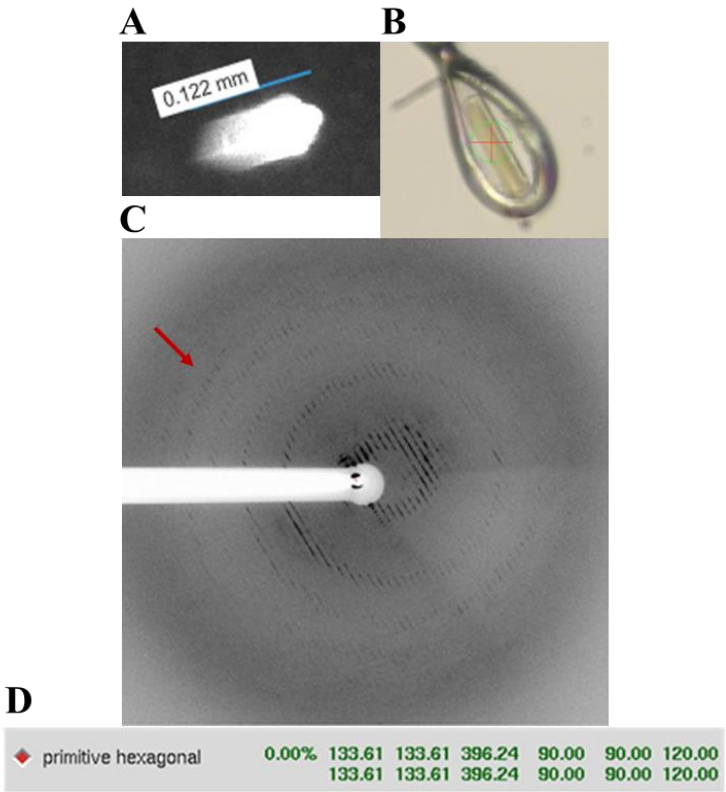


Figure 5.7. Better diffraction of optimized crystals informed that crystals might belong to AcrB contaminant according to its unit cell dimensions and angles. (**A**) Representative optimized crystal. (**B**) Crystal in the nylon loop ready for X-ray diffraction. (**C**) Diffraction of the crystal up to 4.5 Å resolution. (**D**) Space group, unit cell dimensions and angles after indexing. This values match the published AcrB structures.

5.4.3. Protein engineering efforts to improve crystallization of MCM-EE2

5.4.3.1. Introduction of a TEV cleavage site to remove His_6 tag from MCM-EE2

When contaminant proteins like AcrB bind to Ni^{2+}-affinity columns and do not separate by SEC, removal of the hexahistidine (His_6) tag from the purified protein of interest followed by an additional purification step should eliminate the contaminant. We engineered a TEV cleavage site ("ENLYFQS") before the C-terminal His_6 tag of MCM-EE2 (section 5.3.3.3). The purification protocol changed accordingly (**Figure 5.8**); after overnight TEV cleavage, the mixture was applied to the Ni^{2+}-affinity column. Uncleaved MCM-EE2 and AcrB are expected to bind the column whereas the flow-through fractions contain cleaved MCM-EE2 for further purification by gel filtration: the cleaved MCM-EE2 should be contaminant-free after this step. All other protein engineering efforts and crystal optimizations were done with the TEV-cleavable MCM-EE2 variant after AcrB contamination problem came to light.

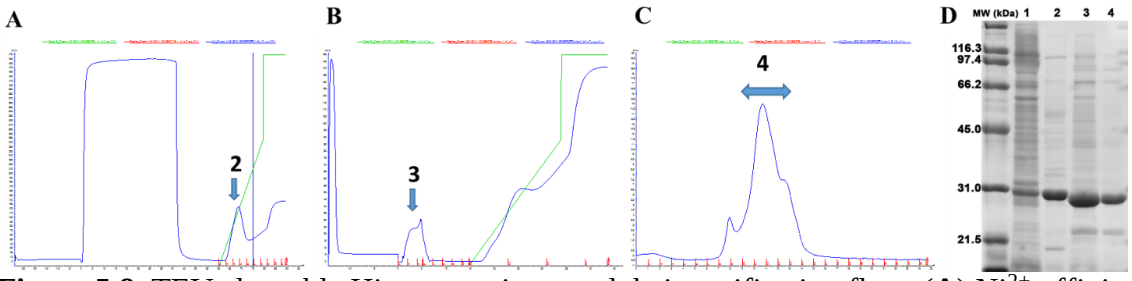


Figure 5.8. TEV cleavable His_6 tag variants and their purification flow. (**A**) Ni^{2+}-affinity purification (buffers as in section 2.2.4) with 0-60% gradient elution. Elution peak (2) is pooled and buffer-exchanged into gel filtration buffer for overnight TEV protease cleavage at 4 °C in 1:1 ratio. (**B**) Another Ni^{2+}-affinity purification (buffers as in section 2.2.4) was run with injection of TEV digested protein, flow-through (3) is concentrated to ~250 µL for SEC. (**C**) SEC with S12 column, second peak (4) is pooled and concentrated for crystallization trials. (**D**) SDS-PAGE of clarified cell lysate (1), Ni^{2+}-affinity (2, 3) and S12 (4) elutions. SPP band shows up under marker band of 31 kDa. TEV cleaved SPP (3, 4) run slightly lower than uncleaved SPP (2).

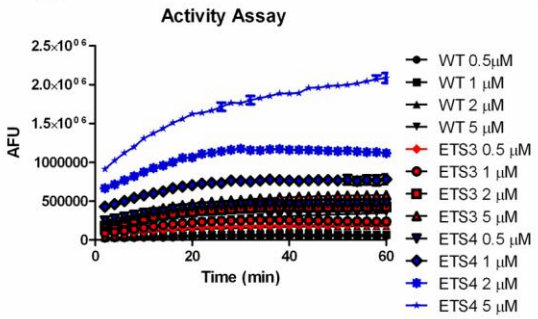
5.4.3.2. Introduction of zinc binding motif to MCM-EE2

On the basis of our observation that rod-shaped crystals only grew in the presence of zinc ions, leading us to think that zinc was required to form a crystal contact, we attempted to engineer a zinc binding motif in the extramembraneous TM6-TM7 loop. We hypothesized that crystallization would be enhanced by installing an explicit zinc binding motif (HEXXH, section 5.3.3.2) to form an additional Zn-dependent crystal contact. However, this alteration did not have any effect on crystallization or quality of diffraction. Later, we inferred that rod-shaped crystals were probably AcrB.

5.4.3.3. Shortening of loop connecting TM6-TM7

An early concern about SPP crystallization was the long loop between TM6-TM7 (section 5.3.3.1). Long and flexible loops are expected to impede crystallization due to heterogeneity in conformation. The TM6-TM7 loop of MCM-EE2 was shortened step-by-step and subjected to sparse matrix crystallization trials. To evaluate whether the loop affects SPP activity, we conducted a diagnostic continuous FRET activity assay with ETS3 and ETS4, and compared activity that of WT MCMSPP (**Figure 5.9**). Notably, the shortened loop variants had higher initial velocities than WT MCMSPP (**Figure 5.9**), though a full Michaelis-Menten analysis was not conducted.

A



B

AFU/min	0.5 μM	1 μM	2 μM	5 μM
WT-MCM	994.7 ± 153.7	1937 ± 172.8	3732 ± 1208	8395 ± 238.2
ETS3	5427 ± 163.2	7801 ± 326.8	10429 ± 230.3	14547 ± 420.3
ETS4	9360 ± 597.5	15668 ± 702.6	22928 ± 599.7	38919 ± 1476

Figure 5.9. Preliminary activity assay for WT-MCM, ETS3 and ETS4. (**A**) FRET assay to determine velocities for different concentrations of enzyme (0.5, 1, 2, 5 μM) with 10 μM Ren390FRET substrate. (**B**) Calculated initial velocities in AFU/min. This data was obtained with one batch of each protein as triplicates in the plate.

Crystallization trials of the variants containing shortened TM6-TM7 loops resulted in new hits with diamond-shaped morphologies (**Table 5.5**). Optimizations proceeded as listed in section 5.3.1. **Figure 5.10** depicts a representative crystal and diffraction limit of ~6 Å resolution. A data set was collected from one of the optimized crystals, but again it turned out to be AcrB which we solved by molecular replacement (*143*). Although we thought that we eliminated the AcrB contamination by using TEV-cleaved constructs, it was probably still in our sample in low amounts which would be enough for its crystallization (*141*). Later, we confirmed presence of AcrB by silver-stained SDS-PAGE gel which has higher sensitivity than that of coomassie-stained (section 5.4.5).

In parallel with the process of shortening the TM6-TM7 loop, we generated another variant with covalently attached lysozyme (mT4L-SPP, section 5.3.3.4). Lysozyme is known to help improve crystallization of membrane proteins due to its hypercrystallizability. A recent paper from Kobilka lab demonstrated the utility of a modified T4 lysozyme in improving their diffraction quality of a GPCR compared to that

of wild-type T4 lysozyme (*138*). Although I screened many different crystallization conditions through sparse matrices, mT4L SPP did not yield any crystals other than those that had morphology consistent with AcrB contaminant crystals.

Table 5.5. New crystal hits with different crystal morphologies after SPP protein engineering efforts.

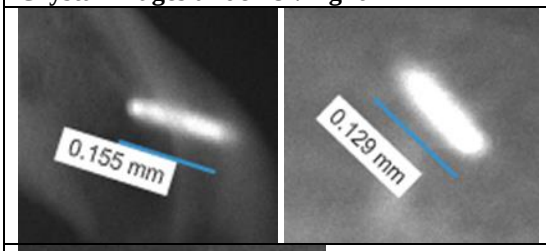
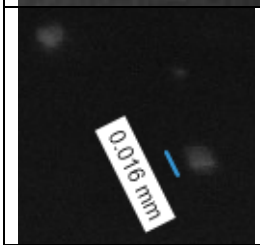
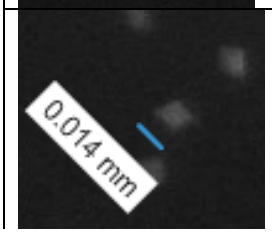
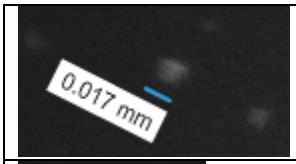
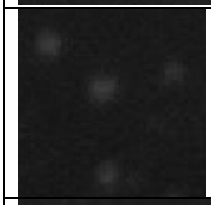
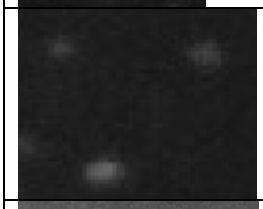
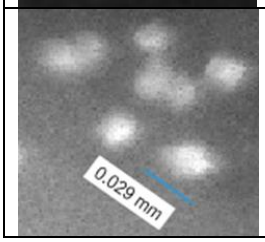
Crystal images under UV light	Variant	Crystallization solution
	MCM-EE2	First crystal hit with rod-shaped morhology 0.1 M NaCacodylate pH 6.5 0.3 M Zn(OAc)2 6% PEG 8000
	ETS1	MemGold 0.04 M Tris pH 8 0.04 M NaCl 27% PEG MME 350
	ETS1	MemGold 0.1 M HEPES pH 7 0.2 M NaCl 22% PEG MME 550
	ETS1	MemGold Sodium Phosphate pH 6.2 0.02 M NaCitrate 18% PEG 2000
	ETS2	MemGold 0.02 M Na Citrate pH 5.6 0.1 M NaCl 11% PEG 3350
	ETS2	MemGold 0.025 M HEPES pH 7.5 0.2 M NaCl 13% PEG 4000

Table 5.5. Continued

	ETS2	MemGold 0.1 M Tris pH 8 0.15 M NaCl 13% PEG 6000
	ETS2	MemGold 0.1 M HEPES pH 7 0.2 M NaCl 22% PEG MME 550
	ETS2	MemGold 0.1 M HEPES pH 7.5 11% PEG 4000
	ETS3	MemGold 0.4 M Ammonium acetate pH 8 13% PEG MME 2000

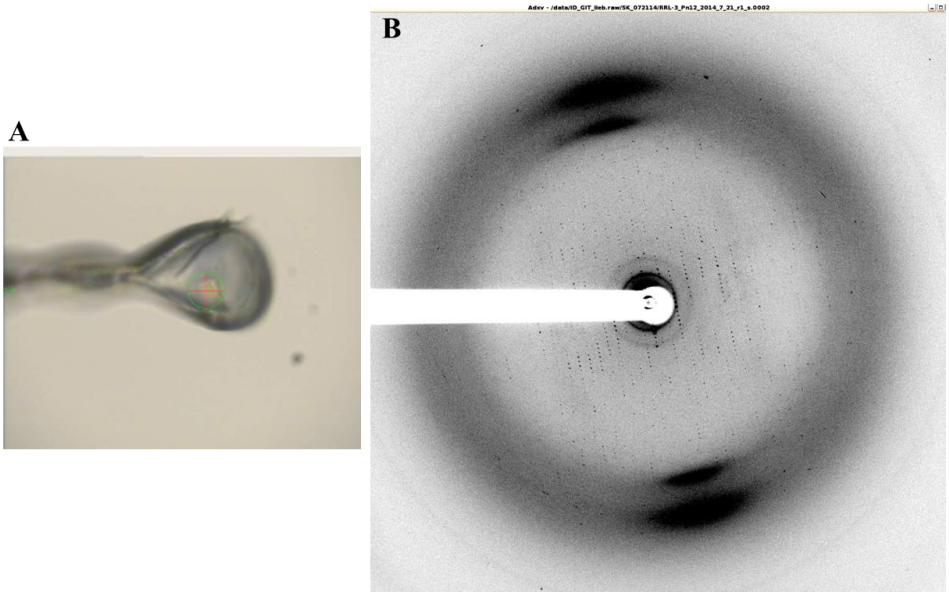


Figure 5.10. Representative diamond-shaped ETS3 crystal diffraction. (**A**) Crystal flash-frozen in a nylon loop for X-ray diffraction. The crystal is grown in solution of 0.4 M Ammonium acetate pH 8, 15% PEG MME 2000 and cryo-protectant is the same solution supplemented with 20% glycerol. 50 µM beam size is represented as green circle on the crystal. (**B**) X-ray diffraction pattern of the crystal. Diffraction spots are observed up to 6 Å resolution. Structure was solved with the collected dataset by molecular replacement using AcrB as a model.

5.4.4. Crystallization trials with bicelles and LCPs

Due to the fact that excess detergent in the crystallization solution can interfere with membrane protein crystallization, alternative crystallization methods were attempted. These alternative methods included DMPC/CHAPSO detergent bicelles, and lipidic cubic phase (LCP) (*135*, *136*, *144*). Both the bicelle and LCP techniques utilize lipids that form bilayers which would mimic natural lipidic environment.

Bicelle crystallization trials with sparse matrix screens were performed for MCM-EE2 and ETS3 variants; neither gave crystal hits. Some of the drops contained precipitation over time (**Figure 5.11**). LCP crystallization trials were performed for MCM-EE2 and cleaved MCM-EE2-TEV variants but again no crystals were identified. **Figure 5.12** shows a representative image of an LCP crystallization tray and one of the drops in the tray.

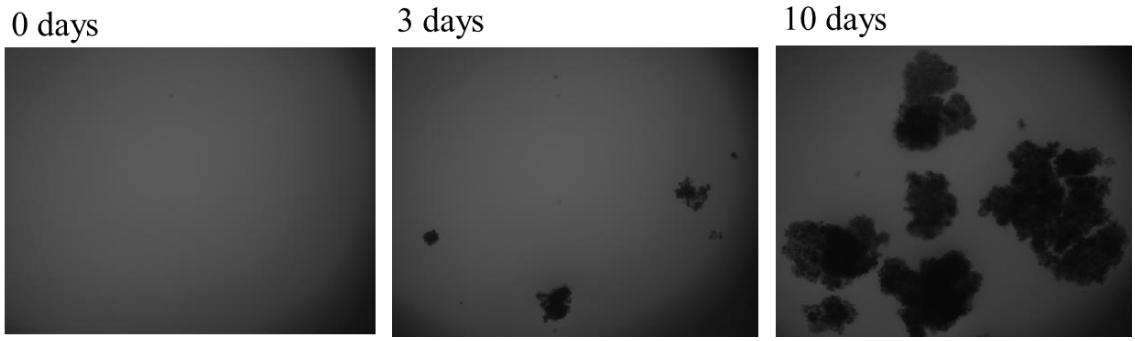


Figure 5.11. Representative images for crystal drops of bicelle crystallization trays over time. It formed aggregates over time.

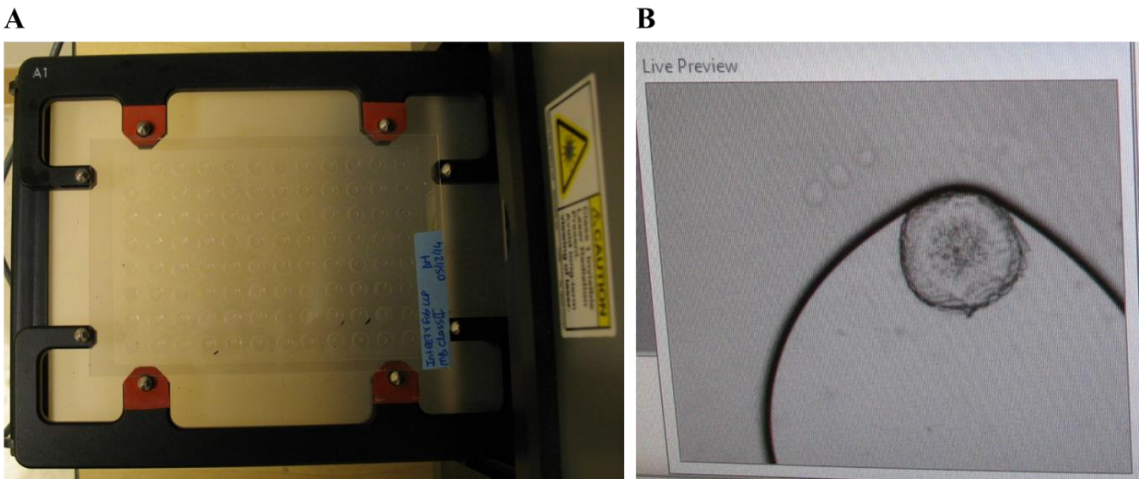


Figure 5.12. LCP trays. (**A**) LCP sandwich plates ready to be imaged by Rigaku Minstrel DT instrument. (**B**) Representative zoomed-in crystal drop in the LCP tray. Protein-monoolein mixture is the small circle inside the crystallization solution from the sparse matrix (big circle).

5.4.5. Confirmation of AcrB contamination

Although almost two years were devoted to optimizing crystallization conditions and SPP constructs, we were unable to fully remove *E. coli* AcrB. Neither removing the hexahistine tag from MCM-EE2 and additional purification steps, nor other protein engineering efforts, proved successful. AcrB was not prominent in SDS-PAGE analysis of purified enzyme when stained by coomassie but AcrB was confirmed with the more sensitive silver-staining protocol when harvested crystals were evaluated by SDS-PAGE analysis (**Figure 5.13**).

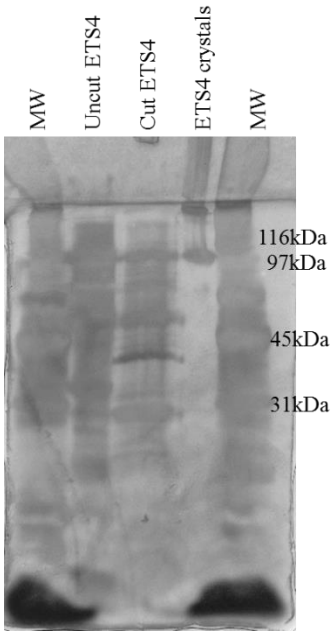


Figure 5.13. Silver-stained SDS-PAGE for ETS4 protein and harvested crystals. 28 rod-shaped ETS4 crystals were harvested and washed 2 times with GF buffer before loading into the SDS-PAGE. When compared to the TEV-cleaved ETS4, crystals give two bands higher in molecular weight than that of ETS4. The lower band (between 97 and 116 kDa band) corresponds to AcrB molecular weight (110 kDa) and the upper band could be the aggregated form of the AcrB protein which could not enter to the resolving gel.

5.4.6. Protein expression in *E. coli BL21(DE3)ΔAcrB*

After facing AcrB crystallization twice, we decided to express our protein in an AcrB knock-out *E. coli* strain*BL21(DE3)ΔAcrB* (*139*), a kind gift from Dr. Yu (Iowa State University). I performed pilot expression trials with both WT-MCM and mT4L-SPP. Yield for cells expression WT-MCM was very low and enzyme could not be purified to any appreciable levels (**Figure 5.14**). At this point, we decided to halt SPP crystallization experiments, and focus our attention on a structure-function study for SPP (Chapter 7).

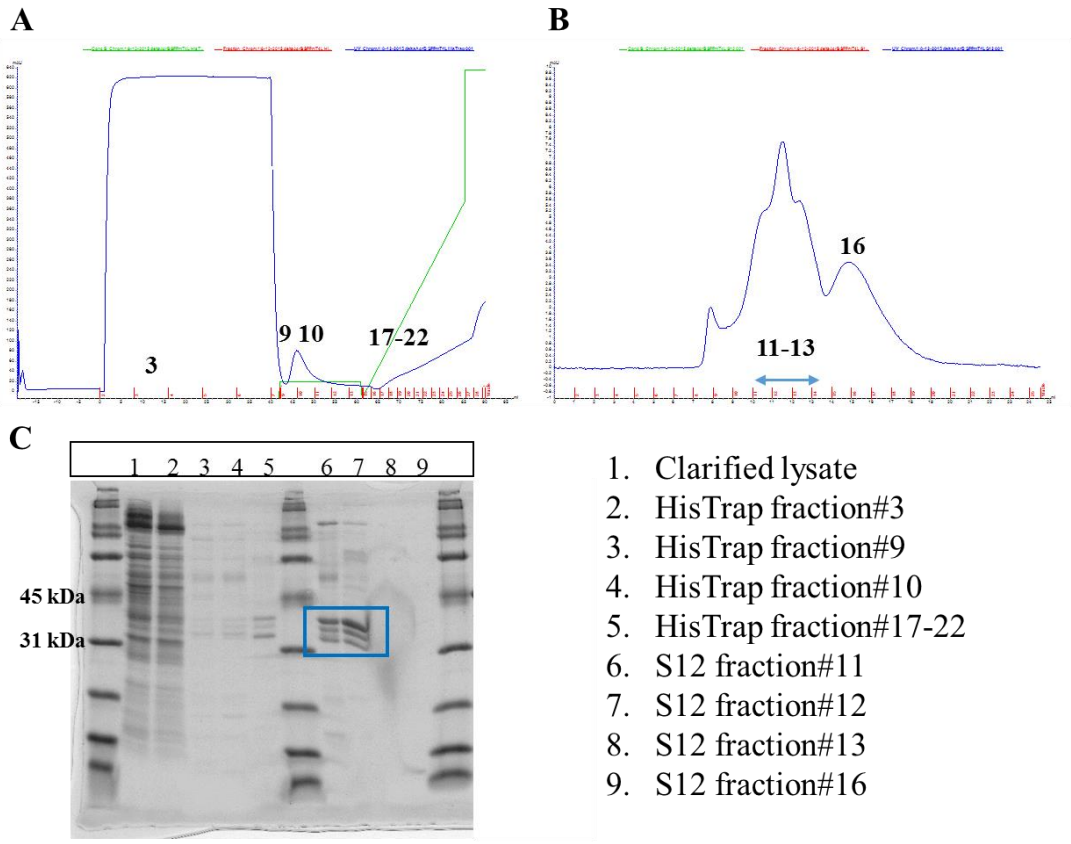


Figure 5.14. Purification flow of WT-MCM expressed in *E.coli BL21(DE3) ΔAcrB*. (**A**) HisTrap purification with 0-60% gradient elution. Elution peak (17-22) are pooled and concentrated to ~250 µL for SEC. (**B**) SEC with S12 column, peak fractions are concentrated for SDS-PAGE. (**C**) SDS-PAGE of cell lysate, HisTrap and S12 fractions. Samples are not pure and three bands show up between 31-45 kDa. All purification buffers contain DDM detergent.

5.4.7. 2D crystallization efforts in collaboration with Schmidt-Krey Lab

In collaboration with Schmidt-Krey Lab, we have been trying to crystallize SPP in 2-dimensional (2D) lipid vesicles. For initial screening, dialysis and negative staining have been performed. I provided 8 batches of WT-MCMΔN23, WT-MCM or WT-MCM-TEV (**Table 5.6**) When we learned that N-terminal 23 amino acids belong to the transmembrane helix 1 of SPP (Chapter 2), we decided to continue all 2D-crystallization experiments with full-length (WT-MCM). We tried to see if SEC column would make

any difference in 2D-crystallization by altering the separation of PDC from empty micelles. Both S12 and S300 columns were tested with WT-MCM and S12 gave better crystals, so we decided to continue with the S12 column. We also tested WT-MCM-TEV, and we concluded that removing C-terminal His_6 tag did not have any discernible effect on 2D-crystallization. Therefore, all experiments were focused on WT-MCM. Early in the project, I had pooled fractions from SEC elution peak for 2D crystallization trials, but an optimization parameter has been keeping fractions separate. There is an overlap between the elution of purified protein and empty micelle; thus each fraction has a different detergent:protein concentration, which could affect crystallization. Indeed, Kasahun discovered a difference in crystal quality depending on fraction (fraction number 11 from S12 column was the best), so fractions have not been pooled in subsequent 2D-crystallization trials. Finally, the purification protocol was changed with 5% buffer B wash after sample loading to eliminate further AcrB contamination (Chapter 2.2.4)

More details on results of dialysis and 2D-crystallization experiments can be found in Schmidt-Krey lab documents: Maureen Metcalfe's master thesis and Kasahun Neselu's laboratory notebooks. Thus far, varying lipid-to-protein ratio, dialysis solution, temperature, etc., has resulted in larger ordered areas (>1 μm) with lower mosaicity (**Figure 5.15**). The next phase of the project will be determining the structure by cryo-Electron Microscopy (cryo-EM).

Table 5.6. Protein samples provided to Schmidt-Krey Lab for WT-MCM 2D-crystallization efforts.

Date	Dialysis and negative-staining by	SPP variant	SEC column and its fractions
12/13/13	Maureen Metcalfe	WT-MCM ΔN23	S300, pooled fractions
05/01/14	Maureen Metcalfe	WT-MCM ΔN23	S300, pooled fractions
07/01/14	Maureen Metcalfe	WT-MCM	S300, pooled fractions
10/08/14	Maureen Metcalfe	WT-MCM	S12, pooled fractions
02/04/15	Maureen Metcalfe	WT-MCM-TEV	S12, pooled fractions
05/13/15	Kasahun Neselu	WT-MCM-TEV	S12, pooled fractions
12/03/15	Kasahun Neselu	WT-MCM	S12, separate fractions
03/14/16	Kasahun Neselu	WT-MCM	S12, separate fractions*

* 5% buffer B wash step was used after sample loading during Ni^{2+} affinity purification

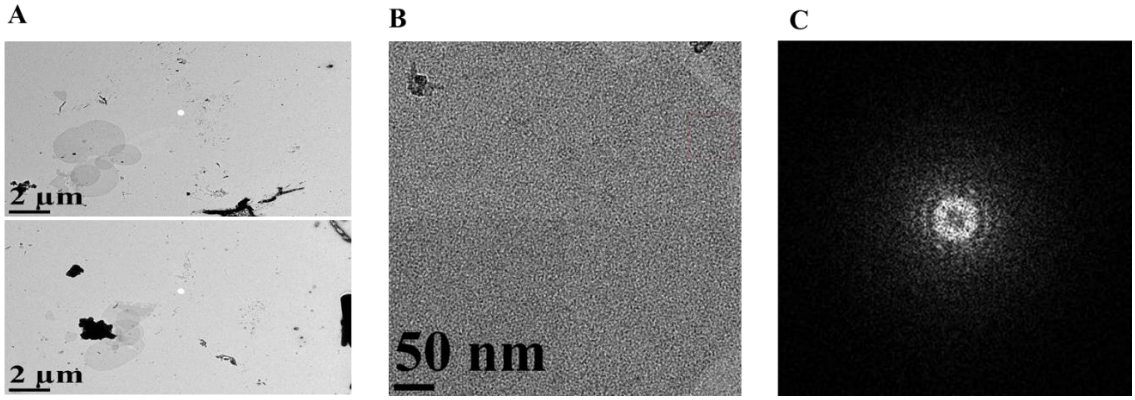


Figure 5.15. Representative 2D-crystallization results of WT-MCM. (**A**) Vesicles formed after 12 days of dialysis. (**B**) Vesicles zoomed-in. (**C**) Live Fast Fourier Transform (FFT) image from the part of the vesicles show good mosaicity.

5.5. Discussion

In this chapter, I cataloged efforts to improve crystallization and diffraction of SPP by optimization of crystallization environment and by protein engineering. I faced AcrB contamination twice and could not eliminate this problem while retaining SPP expression levels required for structure determination. Several other research labs also crystallized AcrB instead of their desired his-tagged membrane protein (*140-142*). AcrB contamination was not evident in the literature until very recently; emerging evidence suggests that ~45% of all membrane protein crystallization trials produce crystals of AcrB even when AcrB is not visible by SDS-PAGE (*141*). *Psakis*, et al. report that AcrB

can crystallize in a variety of different crystallization conditions and suggest that the stress of target protein overexpression might up-regulate AcrB expression. They also propose that crystallization of AcrB in very low quantities (scale of picograms) could happen by its phase-transition of the detergent/lipid content instead of oversaturation due to the crystallization components (*141*).

Many important lessons were learned from this study. During crystallization of any membrane protein of interest, it is always important to check if crystal morphologies are similar to known AcrB crystals (*141*) and if their unit cell dimensions/angles and space groups match with those of AcrB. Even though AcrB is invisible in the SDS-PAGE of the final protein sample, more sensitive methods such as silver-staining should be used to determine the purity of the membrane protein and to identify contaminants like AcrB before valuable time and resources are wasted. The same caution should be taken with 2D-crystallization experiments (*145*). During our 2D-crystallzition efforts, we used to see low amounts of high-contrast regions among vesicles. This was strange because SPP did not have any known extracellular region which would result in crystals with high contrast. Luckily, we did not focus our efforts on these high-contrast vesicles, which we infer by comparison to published data (*145*), was probably AcrB.

During loop shortening studies, we found out that two variants with the shortest TM6-TM7 loop, ETS3 (12 amino acid deletion) and ETS4 (16 amino acid deletion) were catalytically faster than WT-MCM (**Figure 5.9**). ETS4 was even faster than ETS3, this might mean that shorter the loop is, faster the enzyme is. Because the two catalytic aspartates are located in TM6 and TM7, shortening the TM6-TM7 loop may bring the

catalytic aspartates closer to each other, reducing the numbers of inactive conformations accessible to the enzyme.

Because of the challenges at the crystallization level of membrane proteins, research has been done to alter the chemical environment of membrane proteins. Lipid and cell membrane mimics can stabilize membrane proteins and increase the chance of its crystallization, both in 3D and 2D. Lipidic environments such as LCP and bicelles have been developed (*135, 136, 144*) and they have been successfully used in the structure determination of some membrane proteins including many G-protein Coupled Receptor proteins (GPCRs) (*146*). Although their broad utility still needs to be confirmed with more unique membrane protein structures, they hold promise to tackle membrane protein crystallization challenge. More perspectives on membrane protein structure determination are discussed in Chapter 7.

CHAPTER 6: Enzymatic Hydrolysis of 5-nitroanthranilic acid by Metal-dependent Nucleophilic Aromatic Substitution

6.1. Introduction

6.1.1. Nitroaromatic compounds and their biodegradation pathways

A vast number of synthetic and natural organic compounds found in the biosphere (*54, 147*). Nitroaromatic compounds, widely synthesized and used industrially for dyes, pesticides, explosives and solvents, are particularly toxic and recalcitrant to degradation (*148-150*). Nitro substituents are deactivating *meta* directors in electrophilic aromatic substitution reactions, and can form a stable Meisenheimer complex when found *ortho* or *para* to a leaving group in nucleophilic aromatic substitution (S_NAR) (*148-150*). Bacteria from areas contaminated with synthetic nitroaromatics have evolved degradation pathways in which the nitro group is first converted to nitrite, amine, hydroxylamine, keto or hydroxyl by oxygenases or nitroreductases; then ring fission leads to intermediates that can enter central metabolism (*148-150*). Naturally occurring nitro-containing compounds are produced in animals, plants, fungi, and microbes (*150-152*), but documentation of their catabolism is scant (*150*). Discovery of the enzymatic mechanisms for synthesis and biodegradation of such compounds would extend our understanding of metabolic diversity for the prediction and control of their ecological impacts.

There is a growing awareness of the potential use of rhizospheric or endophytic bacteria as biocontrol agents to enhance resistance to plant pathogens (*153-155*), including *Streptomyces scabies* (*156*), which causes potato scab (*157*), a disease of worldwide economic importance (*158*). The primary phytotoxin, thaxtomin, is a family of

5-nitrotryptophan-containing cyclic dipeptides. Among the additional metabolites secreted by *S. scabies* is 5-nitroanthranilic acid (5NAA), the nitrated analog of the tryptophan biosynthesis intermediate chorismate, which is degraded by *Bradyrhizobium sp. JS329*, a bacterial strain isolated from potato field soil (*159*). The ecological role of *Bradyrhizobium sp. JS329* is not currently known, but by eliminating a component of the *S. scabies* chemical arsenal, *Bradyrhizobium sp. JS329* might protect itself and plants from disease (*147, 160*). In contrast to the processes known to biodegrade xenobiotic nitroaromatic compounds, the nitro group is not the initial target for the biodegradation of 5NAA. Instead, 5-nitroanthralinate aminohydrolase (D3WZ85, E.C. 3.5.99.8, 5NAA deaminase, 5NAA-A) performs a unique deamination reaction whereby the amino group, which is *para* to the nitro substituent, is hydrolyzed to form 5-nitrosalicylic acid (5NSA) and the nitro group remains intact. Next, ring fission is catalyzed by 5NSA dioxygenase and the nitro group is eliminated by downstream reactions (*48*) (**Figure 6.1**). Whereas DNA deaminases and other hydrolases can remove amino groups from aliphatic and heterocyclic compounds (*161*), 5NAA-A is the first described hydrolase that can remove an amino group from a substituted benzene.

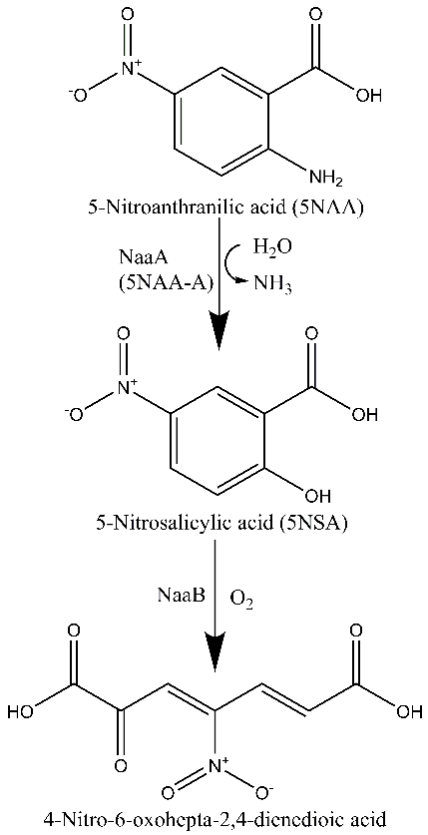


Figure 6.1. Proposed biodegradation mechanism of 5NAA

6.1.2. Chapter overview and publications

The study in this chapter was a collaboration project with Spain Lab in School of Civil and Environmental Engineering. They discovered a new bacterial strain named *Bradyrhizobium sp. JS329* in potato soil and found enzymes encoded its genome which can biodegrade a natural nitroaromatic compound, 5NAA (*48*, *159*). The first step in biodegradation pathway is catalyzed by 5NAA-A, a metalloprotease family member that converts 5NAA to 5NSA (**Figure 6.1**).

Here, we report the crystal structures of this enzyme in the apo, 5NAA-bound, 5NSA/Mn^{2+}-bound 5NAA-A, and Meisenheimer complex intermediate/Zn^{2+}- bound 5NAA-A_{R289A}, as well as characterize metal binding and substrate preferences that tailor

5NAA-A for its unprecedented active site chemistry. We obtained snapshots of its mechanism from its structures in various states. 5NAA-A, an octamer that can use a variety of divalent transition metals for catalysis, employs a novel nucleophilic aromatic substitution mechanism. Unexpectedly, the metal binding site in 5NAA-A is labile. 5NAA-A is specific for 5NAA and cannot hydrolyze other tested derivatives, which are likewise poor inhibitors. The 5NAA-A structure and mechanism expanded our understanding of the chemical ecology of an agriculturally important plant and pathogen, and will inform bioremediation and biocatalytic approaches to mitigate the environmental and ecological impact of nitroanilines and other challenging substrates.

The manuscript of this study was submitted to Nature Chemical Biology in March 22nd, 2016 and it was sent out for review on March 30th, 2016. Mostly positive reviews came back with a couple of requests for additional experiments on May 3rd, 2016. During writing process of this thesis chapter, we did additional experiments, revised our manuscript and submitted revisions to Nature Chemical Biology

6.2. Methods

6.2.1. Molecular biology

5NAA-A was cloned into a pET-30 Xa/LIC vector (Novagen). 5NAA-A variants were generated by site-directed mutagenesis (SDM, QuickChange Lightning Kit, Agilent Technologies) using the 5NAA-A pET-30 Xa/LIC vector as template. Primers were designed with the assistance of PrimerX (Bioinformatics.Org), and synthesized by MWG Operon. Sequences of all the constructs were confirmed by DNA sequencing (MWG Operon). Primer sequences for reported 5NAA-A variants were shown in **Table 6.1**.

Table 6.1. Primers for 5NAA-A variants.

Mutant	Oligonucleotide Primers
H86A	Forward: 5'-AGCCTGTCCTTCAACAGTGCGCTCGACACGATCATGGCG-3'
	Reverse: 5'-CGCCATGATCGTGTCGAGCGCACTGTTGAAGGACAGGCT-3'
E196A	Forward: 5'-GCTCTCGTCGCGGCGGCGACGAACTTC-3'
	Reverse: 5'-GAAGTTCGTCGCCGCCGCGACGAGAGC-3'
E158A	Forward: 5'-CTGCAGTCTGTGGCGCGATTGACTGCGAGCCG-3'
	Reverse: 5'-CGGCTCGCAGTCAATCGCGCCACAGACTGCAG-3'
N124A	Forward: 5'-TCTACGGCTATAGTGTTGTGGCGTGTAAGGGTCCTATGGCGTG-3'
	Reverse: 5'-CACGCCATAGGACCCTTACACGCCACAACACTATAGCCGTAGA-3'
Y223A	Forward: 5'-CCGGGCCGTCCCGCGCGACGCCCTACGTGCC-3'
	Reverse: 5'-GGCACGTAGGGCGTCGCGCGGGACGGCCCGG-3'
Y223F	Forward: 5'-GGCCGTCCCGCTTTACGCCCTACGT-3'
	Reverse: 5'-ACGTAGGGCGTAAAGCGGGACGGCC-3'
R289A	Forward: 5'-GCGGCGTGCCCTACAAGATTTATGCGTTTCCAGAGCTCT-3'
	Reverse: 5'-AGAGCTCTGGAAACGCATAAATCTTGTAGGGCACGCCGC-3'
R373A	Forward: 5'-GGAGTGCAGCATGTGGGCGGACACGAATCCCTACA-3'
	Reverse: 5'-TGTAGGGATTCGTGTCCGCCCACATGCTGCACTCC-3'
D88A	Forward: 5'-TTCAACAGTCATCTCGCGACGATCATGGCGCGCG-3'
	Reverse: 5'-CGCGCGCCATGATCGTCGCGAGATGACTGTTGAA-3'
D88N	Forward: 5'-CTTCAACAGTCATCTCAACACGATCATGGCGCG-3'
	Reverse: 5'-CGCGCCATGATCGTGTTGAGATGACTGTTGAAG-3'
V247K	Forward: 5'-TGCGTATGGCCAAGCTCAAAGAGGCGCTGGAGGAATG-3'
	Reverse: 5'-CATTCCTCCAGCGCCTCTTTGAGCTTGGCCATACGCA-3'

6.2.2. Protein expression and purification

5NAA-A and 5NAA-A variants were expressed in competent *E. coli* Rosetta 2 (DE3) cells. A single transformed colony was inoculated into a selective 20 mL Luria-Bertani culture (LB, Fisher, supplemented with 34 μg/mL chloramphenicol, 50 μg/mL kanamycin) and agitated overnight (16-18 hours) at 37°C, 225 RPM. A starter culture (5 mL) was used to inoculate 1 L of selective LB media and agitated at 37°C, 225 RPM until an optical density at 600 nm of 1.0-1.5 was reached. The cell cultures were cooled with shaking at 225 rpm at 25°C for 1 hour, then induced with 0.5mM isopropyl β-D-thiogalactopyranoside (IPTG, Calbiochem), and allowed to grow overnight (16-18 hours) at 25°C, shaking at 225 RPM. Cells were pelleted (3000 x g, 10 minutes, 4°C) and subsequently flash-frozen with liquid nitrogen and stored at -80°C.

The selenomethionine derivatized enzyme (5NAA-A SeMet) was generated by transforming the pET-30 Xa/LIC plasmid into competent *E. coli* T7 Express Crystal cells (New England Biolabs) following the protocol published by Lambert *et al.* (*162*). Briefly, a single transformed colony was inoculated into 10 mL selective minimal media (5x M9 salts (*163*), 0.4% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.0002% ferric ammonium citrate, 34 μg/mL chloramphenicol, 50 μg/mL kanamycin) containing 50 μg/mL L-methionine. The starter culture was agitated at 225 RPM at 37°C for 2 hours, followed by inoculation at a 1:100 dilution into the selective 1 L minimal media expression media containing 50 μg/mL L-methionine. When an optical density at 600 nm of 0.6-0.8 was reached, cells were pelleted and resuspended into fresh selective minimal media with no L-methionine and shaken at 225 RPM, 37°C, for 2.5 hours. L-selenomethionine was then added to the cell culture to a final concentration of 50 μg/mL; after 30 minutes with agitation, cells were induced with 0.5 mM IPTG, and then allowed to grow overnight (16-18 hours) at 25°C, with 225 RPM agitation. Cells were pelleted by centrifugation (3000 x g, 10 minutes, 4°C) and subsequently flash frozen with liquid nitrogen and stored at -80°C.

Cell pellets were suspended in Ni^{2+}-affinity purification wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 40 mM imidazole, 10% glycerol) containing 0.5x Roche Complete EDTA-free Protease Inhibitor and lysed by two passages through a French Press (13000 psi). Cellular debris was removed by ultra-centrifugation (162,000 x g, 45 minutes, 4 °C). The supernatant loaded onto a 1 mL Ni^{2+}-affinity purification column (GE Healthcare) on an AKTA FPLC system (GE Healthcare), equilibrated with wash buffer. Pure protein was eluted with a gradient from 40-500 mM imidazole by AKTA

FPLC system, elution buffer was exchanged with Factor Xa cleavage buffer (50 mM Tris pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$) using an Amicon 10K MWCO filter (EMD Millipore). The N-terminal hexahistidine tag was removed by incubation with Factor Xa (Roche, 50:1 mass ratio) for 20 hours at 4°C. The digestion mixture was applied to a re-equilibrated 1 mL Ni^{2+}-affinity purification column, the flow-through fractions concentrated in an Amicon 10K MWCO filter, and the concentrate then loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE healthcare) equilibrated with gel filtration buffer (50 mM Bicine pH 7.0, 150 mM NaCl). The gel filtration buffer was selected by assessing 5NAA-A SeMet thermal stability in a range of buffers with various pHs using differential scanning fluorimetry as described previously (*164*). The fractions containing 5NAA-A were concentrated in an Amicon 10K MWCO filter for further use. Protein purity was assessed by 12% SDS-PAGE analysis with Coomassie staining. Protein concentrations were measured by absorbance at 280 nm using molar extinction coefficient (74,461.88 M^{-1} cm^{-1}) determined by total amino acid analysis (Molecular Structure Facility, University of California, Davis) and molecular weight (42,832.31 g/mol). The molecular mass of purified 5NAA-A was estimated by size exclusion chromatography using a Superdex-200 10/300 GL column (GE Healthcare) equilibrated with 5NAA-A gel filtration buffer and calibrated using a gel filtration calibration kit (GE Healthcare). 5NAA-A was analyzed immediately following calibration. The molecular weight of 5NAA-A in solution was determined based on the calibration curve, and the oligomerization state of 5NAA-A was estimated by this molecular weight divided by the calculated mass of a monomer.

6.2.3. Circular dichroism (CD)

CD spectrum and thermal melt were acquired on a Jasco J-810 spectropolarimeter equipped with Neslab RTE 111 circulating water bath and a Jasco PTC-4245/15 temperature control system. CD spectra were taken at room temperature for both gel filtration buffer (blank) and 7.8μM WT 5NAA-A sample in gel filtration buffer. CD thermal melt was performed utilizing a 1 °C/min increase in temperature from 5 to 91 °C. Both CD spectra and thermal melt (every 2 °C) were acquired with 15 averaged scans from 300 to 200 nm at a 200 nm/min scan rate, using a 0.1-cm cuvette. The melting temperature was determined using ellipticity values recorded at 220 nm via sigmoidal dose-response analysis using GraphPad Prism 5 software.

6.2.4. Activity and inhibition assays

To assess the activity of 5NAA-A during purification steps, an analytical high-performance liquid chromatography (HPLC) assay described previously (*159*, *165*) was performed by Zohre Kurt (Spain Lab) on a (a) sample after ultracentrifugation, (b) sample after the first Ni^{2+}-affinity purification (elution), (c) sample after the second Ni^{2+}-affinity purification (flow-through), and (d) sample after gel filtration. 5NAA and 5NSA content was analyzed by HPLC with an Agilent Eclipse XDB-C18 column (4.6 mm by 150 mm; 5 μm) by the method described previously (*159*, *165*).

The chemical structures of substrate 5NAA (Sigma Aldrich) and product 5NSA were confirmed by ^{1}H-NMR (Georgia Tech NMR Center). For characterization of 5NSA, an overnight reaction (0.25 μM 5NAA-A, 1mM 5NAA, 500 μM $Mn(SO_4)$) was performed and 5NSA was purified by HPLC. 5NSA product formation was also

confirmed by mass spectrometry (Georgia Tech Bioanalytical Mass Spectrometry Facility).

A spectrophotometric assay was developed to follow depletion of substrate (λ_{max} for 5NAA= 374 nm; λ_{max} for 5NSA= 306 nm). Divalent transition metals ($MnSO_4$, $Zn(OAc)_2$, $CoCl_2$, $CdCl_2$, $NiSO_4$), $FeSO_4.7H_2O$ (see below) and divalent alkali metals ($MgCl_2$, $CaCl_2$) were screened for 5NAA-A activity. 5NAA (100μM) and varying concentrations of metals (0.25 μM, 2.5 μM, 50 μM, 500 μM) were dispensed in a 96-well plate (Costar 96-Well EIA/RIA). Freshly purified 5NAA-A enzyme (0.25 μM) in gel filtration buffer was added (100 μL total volume) and the reaction mixture was mixed on ice by gentle pipetting. The plate was sealed with optical adhesive film (Micro-Amp) and the reaction was started at 37°C with a Synergy 2 plate reader (BioTek). Absorbance readings were recorded every 3 minutes for 2 hours. 5NAA substrate depletion (374 nm) was used for data analysis. For Michaelis-Menten assays, a range of enzyme concentrations (0.15 μM, 0.25 μM, 0.5 μM) were tested, and 0.25 μM was selected for further assays. Varying concentrations of 5NAA (25-600 μM) and 500 μM $MnCl_2$, $Zn(OAc)_2$, or $CoCl_2$ were distributed and the same procedure above was performed. To minimize oxidation of Fe^{2+} during our assay, the 5NAA-A was first dispensed in a 96-well plate on ice. $FeSO_4{\cdot}7H_2O$ was freshly solubilized as 10mM stock concentration and immediately added (500 μM). The assay was terminated after less than 30 min from the time the Fe^{2+} solution was prepared. For all experiments, absorbance readings of solutions containing all reagents except enzyme were subtracted from those containing enzyme to obtain the signal for 5NAA depletion.

Data were fit using the Michaelis−Menten analysis option within GraphPad Prism 5. The Michaelis−Menten equation is $V = V_{max}[S]/ (K_m + [S])$ where V is the initial velocity, V_{max} is the maximum velocity at infinite substrate concentration, [S] is substrate concentration, and K_m is the substrate concentration at half maximal velocity. Initial velocities were determined by the absorbance reading at 374 nm over the first 15-20 min (negative slope of linear regression). Standard calibration curves were plotted with substrate controls (0-600 μM) for each assay, then initial velocities were plotted against substrate concentrations. Nine replicates were performed for 5NAA-A with $MnCl_2$, $Zn(OAc)_2$, six replicates were performed for 5NAA-A with and $CoCl_2$, $FeSO_4{\cdot}7H_2O$ and six replicates were performed for all 5NAA-A mutants with $MnCl_2$ or $Zn(OAc)_2$. Fe^{2+} data did not fit to a Michaelis-Menten model.

The ability of 5NAA-A to hydrolyze the substrate analogs *m*-nitrobenzoic acid (*m*NBA, Sigma Aldrich), 4-nitroaniline (4NA, Sigma Aldrich), anthranilic acid (AA, Sigma Aldrich), aniline (Alfa Aesar), 2-bromo-5-nitrobenzoic acid (BNBA, Alfa Aesar), 2,4-dinitroaniline (DNAN, Alfa Aesar), aminoterephthalic acid (ATPA, Oxchem), and 4-amino-3-nitrobenzoic acid (ANB, Acros), was monitored by a discontinuous enzyme activity assay. In an eppendorf tube, 0.25 μM 5NAA-A, either 500 μM $MnSO_4$ or $MnCl_2$, and 500-600 μM *m*NBA, 4NA, AA, aniline, BNBA, DNAN, ATPA or ANB were mixed in gel filtration buffer. Substrate 5NAA was used as a positive control. An initial time point and an overnight time point (16-18 hours of incubation in a 37°C water bath) were taken and the reaction was stopped by mixing a sample of the reaction mixture with 10% (v/v) trifluoroacetic acid. The UV-Vis spectra of the reaction mixtures were measured

from 300-500 nm by a BioTek Synergy 2 plate reader to detect any depletion of the respective substrate analog.

Three analogs were tested for inhibition of 5NAA-A. 4NA was dissolved in water, BNBA and ANB were dissolved in 100% (v/v) and 50% (v/v) DMSO respectively. Concentration of purified 5NAA-A (0.25 μM) was kept constant and concentrations of analogs varied from 0.125 μM (0.5x enzyme) to 1000 μM (4000x enzyme). Enzyme, 500 μM $Zn(OAc)_2$ and corresponding analogs were incubated at room temperature for 30-60 min. 5NAA substrate (100 μM) was added and the reaction started as described previously. Only DMSO at highest concentrations of 4NA and BNBA reactions was added in one of the positive controls (5NAA-A with substrate 5NAA) to show that it has no effect on the reaction. Data were fit by the half maximal inhibitory concentration (IC_{50}) within GraphPad Prism 5.

6.2.5. Crystallization

Conditions for (co)-crystallization of 5NAA-A with substrate analogs and/or metal were initially identified by commercial sparse matrix screening (Hampton Index HT and Rigaku Wizard 1 & 2) and further optimized. 5NAA-A SeMet crystals (9.4 mg/mL protein in 50 mM Bicine pH 7.0, 150 mM NaCl) were grown by the sitting-drop method at room temperature in 1.07 M NaH_2PO_4, 0.83 M K_2HPO_4. Crystals of 5NAA-A (7 mg/mL protein in 50 mM Bicine pH 7.0, 150 mM NaCl) with 5NAA (0.22 mM) were grown by the sitting-drop method at room temperature in 1.27 M NaH_2PO_4, 0.63 M K_2HPO_4. Crystals of 5NAA-A (5.8 mg/mL protein in 50 mM Bicine pH 7.0, 150 mM NaCl) with $MnCl_2$ (16 mM) and 5NSA (0.16 mM, Eastman) were grown by sitting-drop method at room temperature in 0.1 M sodium cacodylate pH 6.5, 0.75 M trisodium

citrate. Crystals of 5NAA-A_{R289A} (5.8 mg/mL protein in 50 mM Bicine pH 7.0, 150 mM NaCl) with $Zn(OAc)_2$ (12 mM) and 5NAA (0.22 mM) were grown by sitting-drop method at room temperature from a solution containing 0.1 M sodium cacodylate pH 6.5, 0.65 M trisodium citrate.

6.2.6. Data collection and structure determination

Crystals of 5NAA-A SeMet, 5NAA-A with 5NAA, 5NAA-A with Mn^{2+}/5NSA, and 5NAA-A_{R289A} with Zn^{2+}/5NAA were cryo-cooled in a solution containing the corresponding reservoir solution supplemented with 10% ethylene glycol, 16.6% glycerol, 23% glycerol, and 20.5% glycerol, respectively. Diffraction data were collected at beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne, Illinois. Data were indexed, integrated, scaled and processed using HKL2000 (*166*). The initial structure (5NAA-A SeMet, apo) was solved by multiwavelength Se anomalous phasing using AutoSol(*167*) and the remaining structures (5NAA-A with 5NAA, 5NAA-A with Mn^{2+}/5NSA and 5NAA-A R289A with Zn^{2+}/5NAA) by molecular replacement using Phaser(*168*), using the 5NAA-A SeMet model. The models were iteratively built and refined using Coot(*106*) and Phenix.refine(*107*). eLBOW(*169*) within Phenix was used to generate ligand restraints of 5NAA, 5NSA, and the 5NAA intermediate (Tet, *(R)*-1-carboxy-2-amino-2-hydroxy-5-nitro-cyclohexa-1,4-diene) from coordinates generated using Coot. Readyset (*107*) within Phenix was used to generate metal coordination restraints. With the exception of two loops in apo 5NAA-A (residues 91-108 and 393-397, and see below), polypeptide chains were readily traced from residues 6-425; electron density for the remaining N-terminal residues varied among monomers and structures. 5NSA was modeled in all eight

protomers in the Mn^{2+}/5NSA structure. Tet was only well-resolved in six protomers of 5NAA-A_{R289A} with Zn^{2+}/5NAA. Water molecules are only modeled in the 2.2 Å resolution structure where they were visible. Crystallographic statistics for all four structures were shown in **Table 6.2**.

Table 6.2. Crystallographic data collection and refinement statistics for 5NAA-A structures.

Data Collection	**SeMet Apo**	**5NAA-bound**	**Mn^{2+}/5NSA-bound**	**Zn^{2+}/Tet-bound**
λ (Å)	0.9793	1.0000	1.0000	1.0000
Space group	I422	C2221	C2221	C2221
Mol/ASU	4	8	8	8
a, b, c (Å)	175.27 175.27 354.09	186.46 248.87 249.28	186.77 249.34 249.18	185.88 247.63 247.87
α, β, γ (°)	90 90 90	90 90 90	90 90 90	90 90 90
Resolution (Å)*	78.54 - 2.54 (2.60 - 2.54)	46.28 - 3.22 (3.34 - 3.22)	49.83 - 2.89 (2.99 - 2.89)	37.16 - 2.20 (2.28 - 2.20)
Completeness (%)*	94.83 (71.63)	93.87 (84.91)	92.23 (77.04)	98.31 (98.46)
Wilson B-factor ($Å^2$)	25.35	42.37	51.28	27.80
R_{sym}*	0.11 (0.63)	0.12 (0.58)	0.11 (0.47)	0.10 (0.45)
Redundancy*	10.8 (4.0)	6.7 (5.4)	5.2 (4.7)	5.7 (4.6)
I/σ(I)*	10.41 (1.77)	9.59 (3.59)	11.63 (2.97)	14.63 (3.19)
Refinement				
Resolution (Å)	78.54 - 2.75	46.28 - 3.22	49.83 - 2.89	37.16 - 2.20
No. reflections	86420	87398	119036	281745
R_{work}/R_{free}	0.17/0.21	0.18/0.22	0.18/0.22	0.17/0.20
RMS (bonds) (Å)	0.014	0.013	0.014	0.008
RMS (angles) (°)	1.37	1.35	1.32	1.12
Number of protein atoms	12290	25816	25593	25900
Number of ligand atoms	5	104	112	140
Average B-factor ($Å^2$)	26.0	33.0	39.1	32.3
Ramachandran Analysis				
Favored (%)	96.2	95.8	94.8	96.3
Allowed (%)	3.6	3.9	4.6	3.2
Outliers (%)	0.2	0.3	0.6	0.4

* Highest resolution shell values are in the parenthesis

6.2.7. Isothermal titration calorimetry (ITC) measurements

For ITC data shown, 5NAA-A was prepared in gel filtration buffer at 6.4 μM in 1.8 ml aliquots. Stock solutions of 1 mM 5NAA, 1 mM 5NSA, 3.5 mM ANB, 3.2 mM 4NA, 100 mM $Zn(OAc)_2$ and 100 mM $MnCl_2$ were diluted to 0.15 mM with gel filtration buffer accordingly. All samples were degassed extensively on a vacuum line (MicroCal

ThermoVac) before ITC measurements (MicroCal VP-ITC). In a typical experiment, an automated sequence of 30 injections of 9 μL of sample was applied to the protein in the sample cell with 6 min between injections to allow for equilibration. The reaction solution was stirred at 270 RPM, and the temperature of the chamber was maintained at 25°C. Several ITC parameters were varied in an attempt to detect a heat of binding of metal ion to 5NAA-A, including changing the enzyme concentration (7.8 μM), metal ion concentration (250, 500 μM), salt (e.g. $ZnCl_2$), and for the experiment in which 5NAA-A was preincubated with 5NAA, time of incubation (1 h).

Data were analyzed with the Origin 7.0 software package from MicroCal, using one-site binding model. A background correction was applied to each experiment by subtracting the average of the last four injections. Background corrections were in accordance with control experiments (gel filtration buffer). A non-linear least squares method was used to obtain one-site model best fit parameters for the number of binding sites, n, the association constant, K_a, and the change in enthalpy, $\Delta H°$. The dissociation constant, K_d was calculated by taking the inverse of K_a.

6.2.8. Total reflection x-ray fluorescence (TXRF) measurements

TXRF measurements were performed with a Bruker S2 Picofox TXRF (Bruker Instruments, Inc.). Quartz glass disks were used as sample carriers (Bruker Instruments, Inc.). The excitation settings were 50 kV and 750 mA. Measurements were performed by signal integration over 1000 seconds. Quantification of the analytes was performed by Bruker S2 Picofox TXRF software. After overnight dialysis of apo 5NAA-A against either Mn^{2+} or Zn^{2+} (50x excess of protein concentration) by 10K Slide-A-Lyzer Dialysis Cassette, the protein solution was re-run on Superdex-200 10/300 GL column with gel

filtration buffer to remove excess metal. Such experiments were performed twice, and quantification was based on the known quantity of the internal standard gallium (1 ppm) and amount of protein (5NAA-A) to calculate metal:protein ratio. For determination of manganese and zinc amount in *E.coli* and *Bradyrhizobium sp. JS329* cell lysates, different concentrations of $MnCl_2$ and $ZnCl_2$ (0.1, 10, 1000 μM) were analyzed with TXRF along with samples to plot standard curve.

6.2.9. Size-exclusion chromatography - small angle x-ray scattering (SEC-SAXS)

SEC-SAXS experiments were performed and analyzed by Srinivas Chakravarthy at BioCAT (beamline 18ID, Advanced Photon Source at Argonne National Labs). The basic setup for SEC-SAXS was based on the pioneering experiment of this kind (*170*). The camera included a focused 12 keV (1.03 Å) X-ray beam, a 1.5 mm quartz capillary sample cell, a sample to detector distance of ~3.5 m, and a Pilatus 3 1M detector (Dectris). The q-range sampled was ~0.0042-0.38 $Å^{-1}$. In order to ensure sample monodispersity, we used an in-line SEC setup, which included an AKTA FPLC system with a Superdex-200 Increase 10/300 GL column (GE Healthcare). The elution trajectory after the UV monitor was redirected to the SAXS sample flow-cell. One second exposures were collected every 2 seconds during the gel-filtration chromatography run. Appropriate exposures were averaged and used as the buffer curve for each run, and the exposures during elution (co-incident with the UV peak on the chromatogram) were treated as protein+buffer curves. Data were corrected for background scattering by subtracting the buffer curve from protein+buffer curves. Radius of Gyration (R_g) was calculated using Guinier approximation and a P(R) curve, both of which were done using

PRIMUS (*171*). D_{max} was also calculated using the P(R) curve and the output was then used to calculate molecular envelopes using DAMMIN 9 separate times (*172*). These envelopes were then averaged using DAMAVER and DAMFILT (*173*). The high resolution crystal structure was fit into the envelope using SUPCOMB (*174*).

6.2.10. In silico sequence and structure analysis

Sequence alignments were conducted using CLUSTAL-Omega (*175*). Comparison to known structures in the PDB was performed using BioXGEM (*176*, *177*). Interfaces were calculated with PDBePISA (*105*). Tunnels and channels were evaluated in CAVER (*178*), and homology models were generated in SWISS-MODEL (*179*). Evolutionary analyses were conducted in MEGA6 (*180*). The 5NAA-A sequence was queried against the non-redundant protein sequence database, allowing for up to 20,000 target sequences. The resultant 2,256 sequences were then aligned using MUSCLE (*181*). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (*182*) and the tree with the highest log likelihood (-372632.2194) is presented. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 268 positions in the final dataset. All structure figures were generated in PyMOL (www.pymol.org). Unless stated otherwise, interaction distances are reported for a representative protomer.

6.3. Results

6.3.1. Overall structure

The crystal structure of apo 5NAA-A (Selenomethionine (SeMet) derivatized apo, 2.75 Å resolution, **Table 6.2**), confirms inferences from sequence that 5NAA-A belongs to the MEROPS M20 metallopeptidase family (*50*) and thus evolved from a common ancestor (*51*). The closest structural homologs are M20 family members: di-zinc carboxypeptidase G2 (root mean squared deviation (R.M.S.D.): 5.04 Å, (sequence similarity ~35%, **Figure 6.2**) (*16*), N-acetyl-L-citrulline deacetylase (ACD, R.M.S.D.: 4.32 Å) (*52*), and N-succinyl-L,L-diaminopimelic acid desuccinylase (SDAD, R.M.S.D.: 5.66 Å) (*53*) (**Figure 6.3**), but 5NAA-A harbors key distinctions leading to its unique supramolecular arrangement and catalytic site (see below).

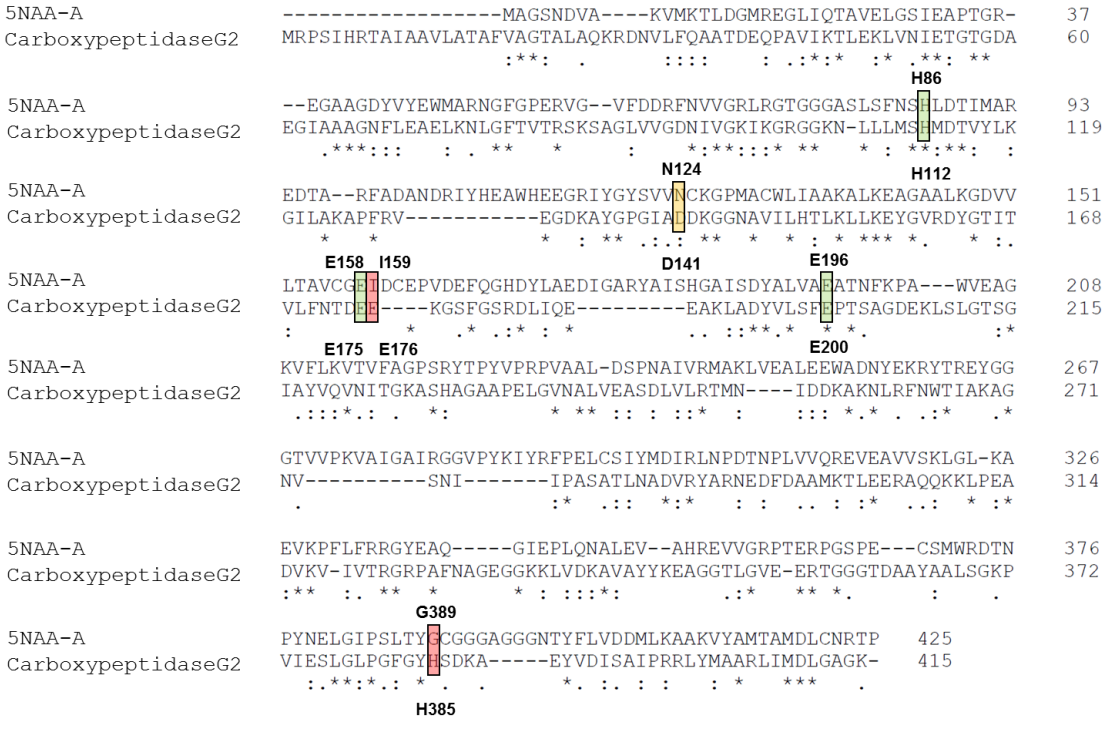
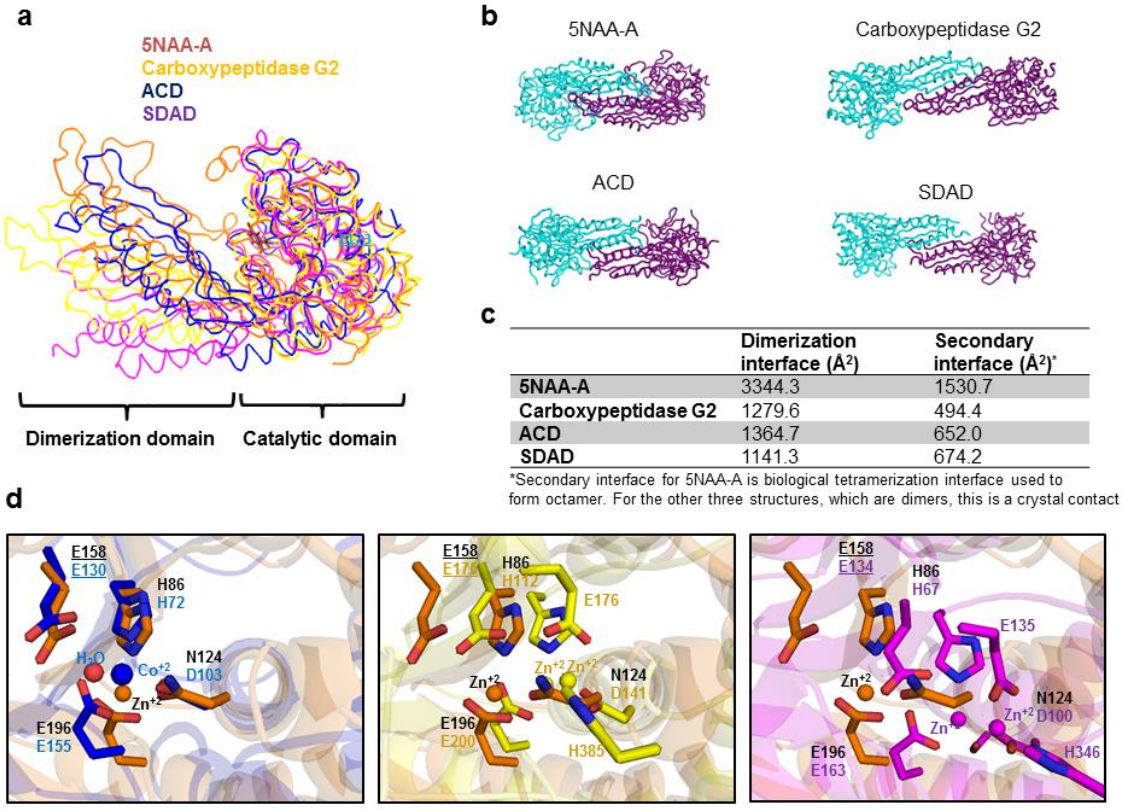


Figure 6.2. Sequence alignment of 5NAA-A with carboxypeptidase G2 (PDB ID: 1CG2). Metal binding/active site residues are highlighted and labeled on top for 5NAA-

A, bottom for carboxypeptidase G2. Conserved residues: * or green highlight; similar residues: :, ., or yellow highlight; non-conserved active-site residues highlighted in red.



	Dimerization interface (Å²)	Secondary interface (Å²)*
5NAA-A	3344.3	1530.7
Carboxypeptidase G2	1279.6	494.4
ACD	1364.7	652.0
SDAD	1141.3	674.2

*Secondary interface for 5NAA-A is biological tetramerization interface used to form octamer. For the other three structures, which are dimers, this is a crystal contact

Figure 6.3. Structural comparison of 5NAA-A and M20 family members carboxypeptidase G2 (PDB ID: 1CG2), N-acetyl-L-citrulline deacetylase (ACD, PDB ID: 2F7V), and N-succinyl-L,L-diaminopimelic acid desuccinylase (SDAD, PDB ID: 3IC1). (**a**) Superposition of monomers. (**b**) Dimer orientation comparison. (**c**) Calculated interface areas of 5NAA-A, carboxypeptidase G2, ACD, and SDAD. (**d**) Pairwise superposition of metal center residues of 5NAA-A (orange) with (left) ACD (blue), (middle) carboxypeptidase G2 (yellow), and (right) SDAD (magenta). Residue labels are in black for 5NAA-A and in annotated colors for the others; catalytic glutamates are underlined.

Analysis of the crystal lattice and small angle X-ray scattering envelope reveal an octamer in which the eight ~43 kDa 5NAA-A protomers are arranged radially around a ~50 Å diameter solvent channel (**Figure 6.4a, b** and **Figure 6.5**) to form a ~350 kDa enzyme. The octamer is unique among characterized M20 family members, which are at most functional dimers (*183*). A shift in the relative orientation of the two lobes, labeled

"catalytic" and "dimerization" (*16*), and novel loop configurations, generate two large surface areas (**Figure 6.3a, b, c**). The first "dimerization" interface establishes the 5NAA-A catalytic center (see below) and is larger than that of carboxypeptidase, ACD, or SDAD (**Figure 6.4a** and **Figure 6.3c**). The importance of the dimerization interface is underscored by the fact that the V247K variant, designed to destabilize this interface, could not be purified to appreciable levels (**Table 6.3**). The second "catalytic" interface enables the closed circular arrangement of a tetramer of dimers (**Figure 6.4a**).

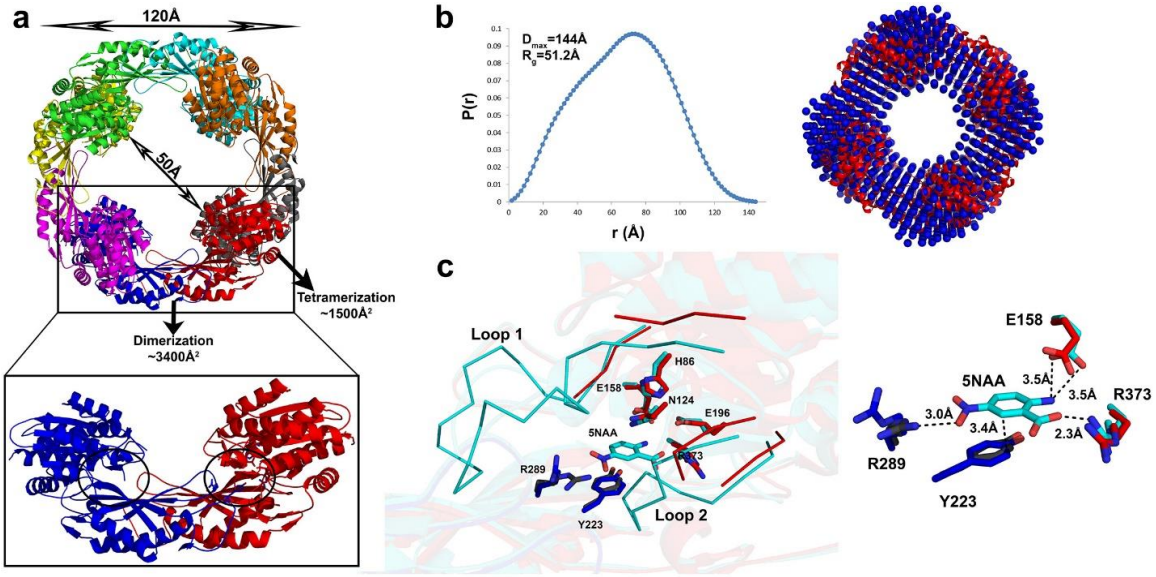


Figure 6.4. Overall structure and substrate binding of 5NAA-A. (**a**) Octameric crystal structure of 5NAA-A, colored by protomer (apo structure shown). Overall dimensions, dimerization, and tetramerization interfaces are labeled. The two active sites within a dimer are circled. (**b**) Solution structure of 5NAA-A confirms donut shaped octamer. Left: pair distribution P(r) plot with calculated maximum particle size (Dmax) and radius of gyration (Rg). Right: SAXS envelope (blue) superimposed with the apo crystal structure (red). (**c**) Substrate binding details. Left: Superposition of apo and 5NAA-bound structures in binding pocket. Loop1 (residues 91-108) and Loop 2 (residues 393-397) are disordered in the apo structure (red/dark blue) but adopt an ordered conformation when 5NAA is bound (cyan/grey). Right: 5NAA binding interactions with protein residues, superimposed with apo structure.

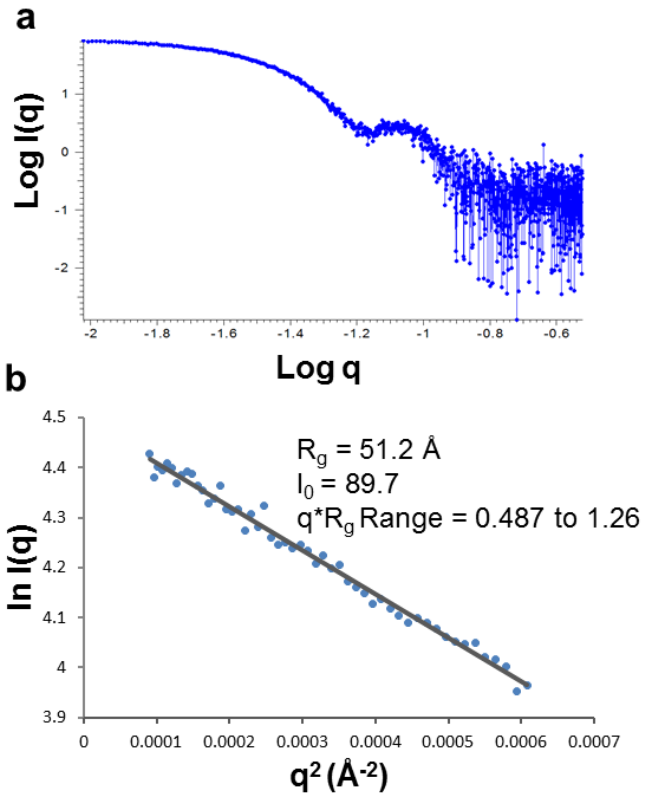


Figure 6.5. Small Angle X-ray Scattering (SAXS) analysis. (**a**) Scattering profile. (**b**) Guinier plot. I(q): scattered intensity, q: scattering vector.

Table 6.3. 5NAA-A mutants, their expression levels, and residual activities.

Residue	**Mutation**	**Activity Test**
Metal binding	H86A	Soluble expression, inactive
	N124A	No soluble expression
	E196A	Soluble expression, inactive
Substrate binding	R289A	Soluble expression, inactive
	R373A	Soluble expression, inactive
	Y223A	Soluble expression, inactive
	Y223F	No soluble expression
Catalysis	E158A	Soluble expression, inactive
Second shell	D88A	Precipitation after purification
	D88N	Soluble expression, inactive
Dimerization	V247K	No soluble expression

6.3.2. Substrate binding

The co-crystal structure of 5NAA-A with bound substrate (5NAA-bound, 3.2 Å resolution, **Table 6.2**) reveals that 5NAA binds within one catalytic lobe but stabilized by a key residue Tyr_{223}, from an adjacent monomer, which provides parallel-displaced π-π stacking stabilization (**Figure 6.4c**). Within the catalytic lobe of one monomer, 5NAA is held in place by Arg_{373} and Arg_{289}, which stabilize the carboxyl and nitro groups, respectively, and the side chain of Glu_{158}, which is within hydrogen bonding distance of the amino substituent. Asp_{160} and the main chain of Trp_{372} also stabilize the aforementioned adjacent Tyr_{223} (**Figure 6.6**). Alanine variants of Tyr_{223}, Arg_{289}, and Arg_{373}, as well as Y223F, are all inactive or not isolated in sufficient yield for further characterization (**Table 6.3**). Two loops disordered in the apo structure are located in this region of the catalytic lobe. Substrate binding, even in the absence of metal, organizes these loops, primarily by interactions between Tyr_{288} and Arg_{98} and between Asp_{95} and Lys_{286} (**Figure 6.4c** and **Figure 6.6b**).

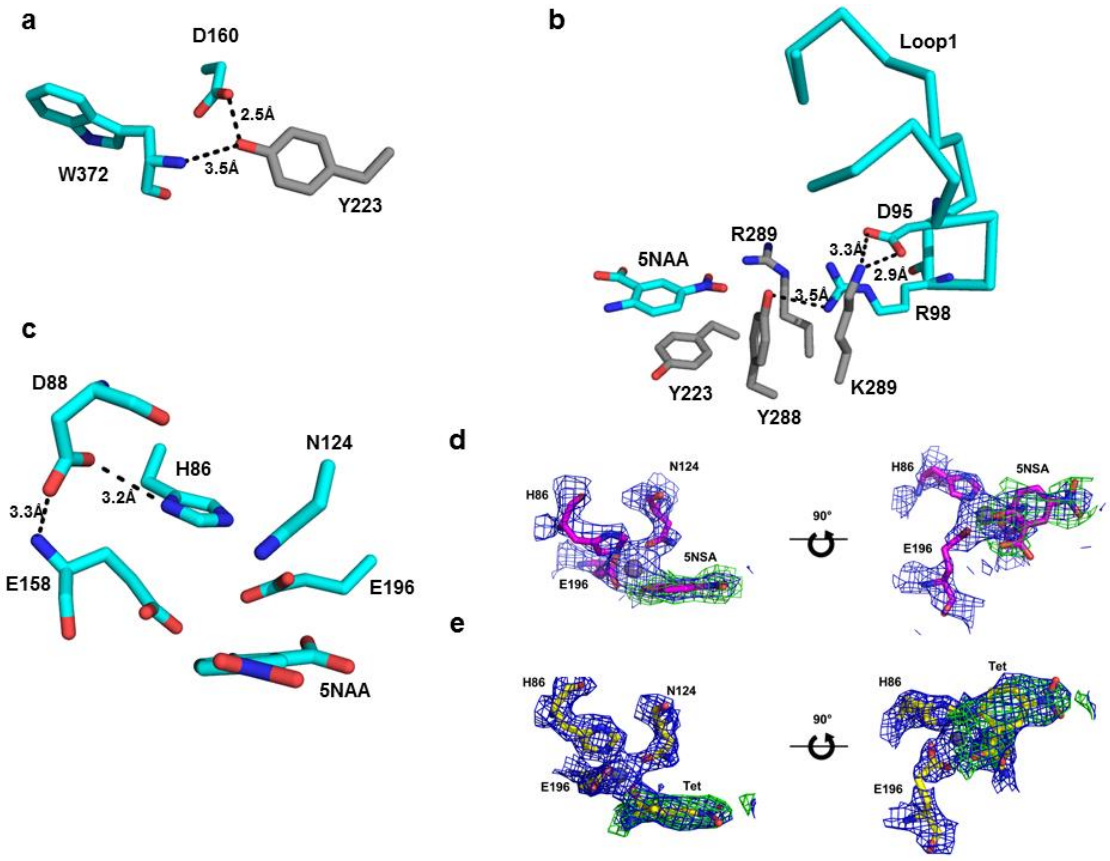


Figure 6.6. Selected stabilizing interactions and representative electron density. Stabilizing interactions of 5NAA-bound structure; (**a**) Y223 (**b**) Loop1 (**c**) Active site second shell interactions. Cyan and grey residues derive from different monomers in the dimer. The same interactions were observed for Mn^{2+}/5NSA- and Zn^{2+}/Tet- bound structures (not shown). Representative electron densities for Mn^{2+}/5NSA- and Zn^{2+}/Tet-bound structures (**d**) Metal binding residues and 5NSA in the active site of Mn^{2+}/5NSA-bound structure (**e**) Metal binding residues and Tet in the active site of Zn^{2+}/Tet-bound structure. Final 2Fo − Fc electron density (blue) is contoured at 2.0σ and difference Fo − Fc density (green) is contoured at 2.5σ. The Fo − Fc density map was calculated based on coordinates of the refined structure refined prior to modeling 5NSA/Tet.

6.3.3. Metal binding

Adjacent to the 5NAA binding site is a triad of ligands (His_{86}-Glu_{196}-Asn_{124}), in a location equivalent to that of the corresponding M20 metalloprotease His_2-Glu_2-Asp metal binding motif. 5NAA-A requires metal ions for hydrolysis (**Figure 6.7a, b** and **Figure 6.8a**) (*48*). Coordinating ligands His_{86} and Glu_{196} are oriented identically as in carboxypeptidase, ACD, and SDAD (**Figure 6.3d**). Asn_{124}, which replaces the Asp found

in dinuclear M20 metalloproteases, cannot act as a bridging ligand. At the position of the two remaining ligands that form the second metal binding site in these dinuclear enzymes reside residues Ile_{159} and Gly_{389} in 5NAA-A (**Figure 6.2**). Neighboring second shell residues such as Asp_{88} and Glu_{158} provide hydrogen-bonding stabilization to orient metal binding ligands even in the absence of metal (**Figure 6.6c**). Alanine mutants of His_{86}, Glu_{196}, Asn_{124}, Asp_{88}, Glu_{158}, as well as D88N, are not active or result in a misfolded protein that cannot be purified for further characterization (**Table 6.3**). With 5NAA bound, there are no obvious tunnels or channels leading to the metal center, and all attempts to obtain a structure of holo 5NAA-A have been unsuccessful.

Several lines of biochemical evidence support a mononuclear, labile, low affinity, and promiscuous metal binding site. Rates with Mn^{2+} and Zn^{2+} were faster than other metal ions tested (**Figure 6.7b** and **Figure 6.8a**), but only in the presence of unexpectedly high excess of metal over enzyme (**Figure 6.7a**). Despite optimizing numerous parameters, we could not detect heat of binding for either metal ion to 5NAA-A by ITC (Methods, **Figure 6.7c** and **Figure 6.9a**). However, titrating 5NAA alone, or a mixture of 5NSA and Mn^{2+}, into 5NAA-A reveals a binding dissociation constant (K_d) in the low micromolar range, tighter than that obtained for titrating 5NSA alone (**Figure 6.7d**) or a mixture of 5NSA and Zn^{2+} (**Figure 6.7e** and **Figure 6.9a**). In addition, no heat of binding was detected when Mn^{2+} was titrated into a solution of 5NAA-A pre-incubated with 5NAA (**Figure 6.7f**). Finally, even though high activity is measured from *E. coli* clarified cell lysate containing wild-type recombinant 5NAA-A (7.7 nmol/min/mg using an end-point assay (*159*)), all activity is lost at the first affinity chromatography step at

which point the enzyme is apo, and cannot be reconstituted with either Zn^{2+} or Mn^{2+} (**Figure 6.8b**).

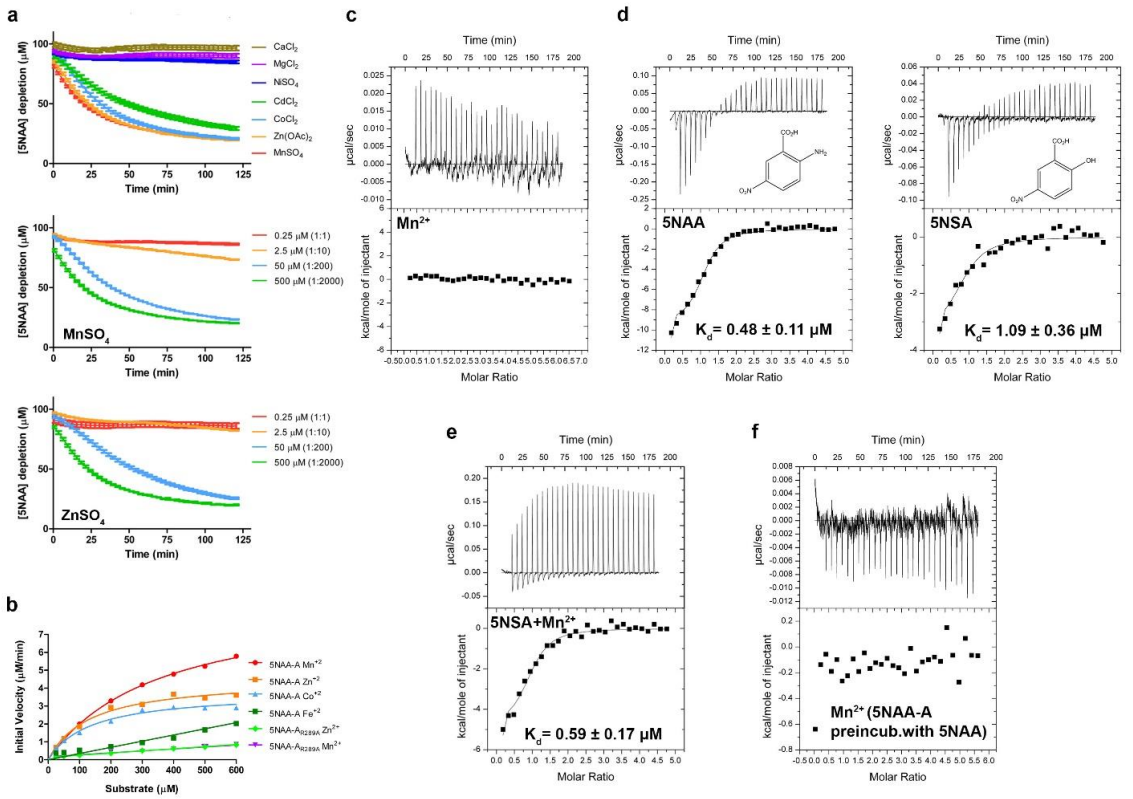
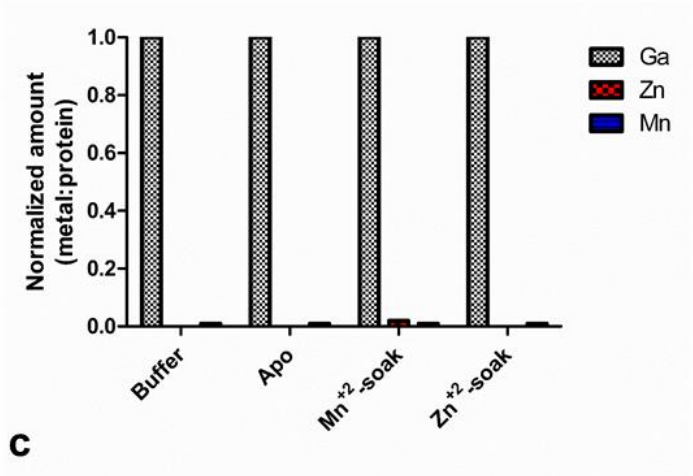


Figure 6.7. Biochemical characterization of 5NAA-A and metal dependence of activity. (**a**) Top: Effect of divalent transition or alkaline earth metals on 5NAA-A activity as measured by 5NAA depletion (500 μM metal ion, 0.25 μM 5NAA-A). Concentration dependence of Mn^{2+} (middle panel) and Zn^{2+} (bottom panel) on 5NAA-A activity. Triplicate runs are performed for each metal and its concentration (mean ± standard error of the mean). (**b**) Kinetics analysis for 5NAA-A in the presence of activating metal ions and representative mutant 5NAA-A_{R289A}. Three triplicate runs are performed for Mn^{2+}, Zn^{2+} analysis and two triplicate runs are performed for Co^{2+}, Fe^{2+}, Zn^{2+} and Mn^{2+} with 5NAA-A_{R289A} analysis (mean ± standard error of the mean). (**c-f**), Metal ion, substrate, and product binding measured by isothermal titration calorimetry. Corresponding plots of integrated heat versus molar ratio and calculated K_ds are shown below raw data (see also **Figure 6.9**). For each experiment, 150 μM of titrant indicated was injected in 9 μl aliquots to 6.4 μM enzyme except in (f), where 5NAA-A was pre-incubated with 150 μM 5NAA at room temperature for 30 min prior to injection of 150 μM Mn^{2+} aliquots. Representative thermograms for each ITC experiment are shown. Each run is repeated three times except pre-incubation run which is repeated two times. Parameters shown inset are calculated from triplicate analysis (mean ± standard deviation).

a

	K_M (μM)	k_{cat} (s^{-1})	(k_{cat}/K_M)x10^3 ($M^{-1}s^{-1}$)
Zn^{2+}	133.0 ± 12.9	0.302 ± 0.009	2.271 ± 0.101
Co^{2+}	139.6 ± 9.8	0.256 ± 0.006	1.834 ± 0.074
Mn^{2+}	354.0 ± 14.2	0.605 ± 0.012	1.709 ± 0.045

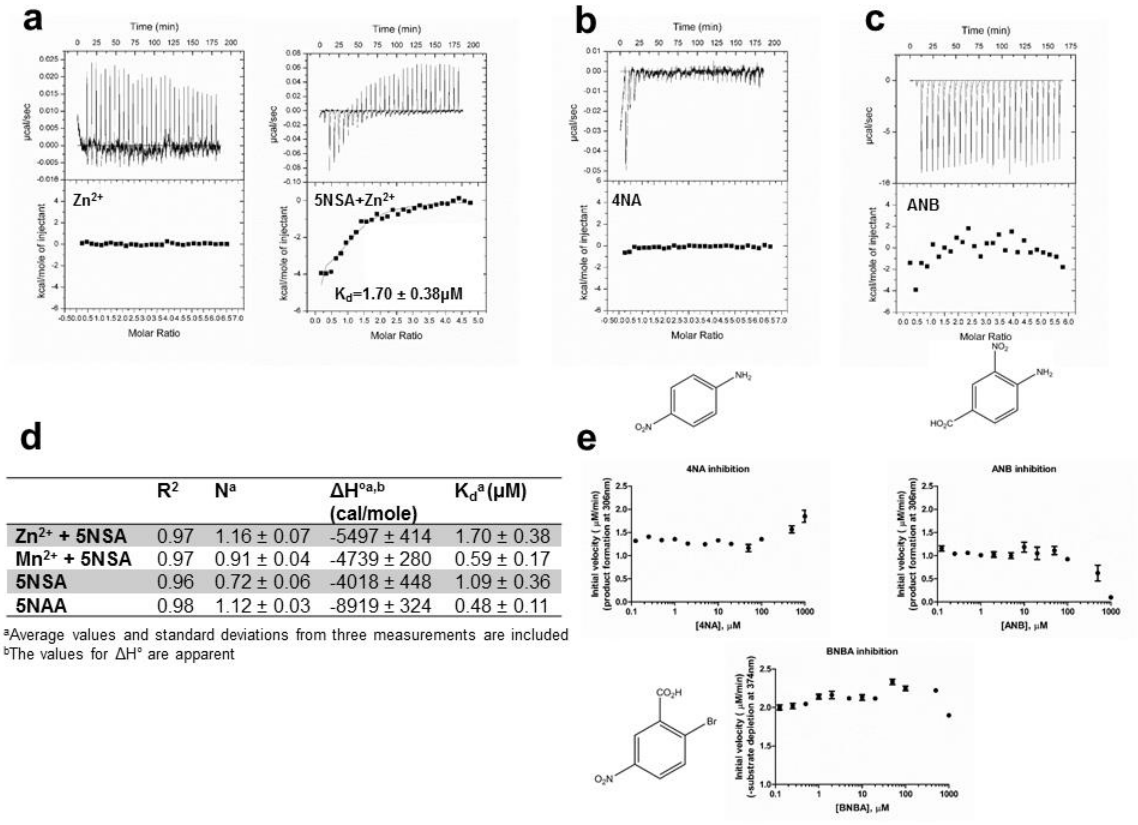
b



c

	Mn (μM)	Zn (μM)
E.coli (DE3)	5.3	65.2
Bradyrhizobium sp. JS329	5.1	56.1

Figure 6.8. Metal content and dependence on enzyme activity. (**a**) Kinetic parameters of 5NAA-A with Mn^{2+}, Zn^{2+} and Co^{2+}. Three triplicate runs are performed for Mn^{2+}, Zn^{2+} analysis and two triplicate runs are performed for Co^{2+} analysis (mean ± standard deviation). (**b**) Total X-ray Fluorescence (TXRF) data to detect bound metal after overnight dialysis of apo 5NAA-A against 50x excess Mn^{2+} or Zn^{2+}, followed by gel filtration. Data for buffer only and apo prior to dialysis are included. The experiment is performed twice and a representative result is shown. (**c**) TXRF analysis of manganese and zinc contents of *E.coli (DE3)* and *Bradyrhizobium sp. JS329* clarified cell lysates. The experiment is performed twice and mean result is shown.



d

	R^2	N[a]	$\Delta H^{o a,b}$ (cal/mole)	K_d[a] (μM)
Zn^{2+} + 5NSA	0.97	1.16 ± 0.07	-5497 ± 414	1.70 ± 0.38
Mn^{2+} + 5NSA	0.97	0.91 ± 0.04	-4739 ± 280	0.59 ± 0.17
5NSA	0.96	0.72 ± 0.06	-4018 ± 448	1.09 ± 0.36
5NAA	0.98	1.12 ± 0.03	-8919 ± 324	0.48 ± 0.11

[a]Average values and standard deviations from three measurements are included
[b]The values for ΔH° are apparent

Figure 6.9. Additional ITC binding data and enzyme inhibition profiles for substrate analogs. ITC data; (**a**) Left: Zn^{2+} only; Right, pre-mixed 5NSA and Zn^{2+}. (**b**) 4NA. (**c**) ANB. Raw data and corresponding plot of integrated heat versus molar ratio are presented. In each experiment, 6.4μM 5NAA-A was titrated with aliquots of 150 μM ligand(s). Representative thermograms for each ITC experiment are shown. Runs of Zn^{2+} and 5NSA+Zn^{2+} are repeated three times, runs of 4NA and ANB are repeated two times. (**d**) Thermodynamic parameters for ITC experiments with one-site best fit model. Parameters are calculated from triplicate analysis (mean ± standard deviation). (**e**) Inhibition profiles of WT 5NAA-A with substrate analogs. Half maximal inhibitory concentration (IC_{50}) plots and chemical structures of 4NA, ANB, and BNBA.

Consistent with biochemical results, we co-crystallized Mn^{2+} with the product 5NSA and wild-type 5NAA-A (Mn^{2+}/5NSA-bound, 2.9 Å resolution), and Zn^{2+}, and 5NAA with the inactive mutant 5NAA-A_{R289A} (Zn^{2+}/Tet-bound, 2.2 Å resolution), which trapped the Meisenheimer intermediate (Tet, (*R*)-1-carboxy-2-amino-2-hydroxy-5-nitrocyclohex-1,4-diene, see below) (**Table 6.2** and **Figure 6.6d, e**). The two structures confirm the His_{86}-Glu_{196}-Asn_{124} metal ligand residues. The structures superimpose well

with that of 5NAA-bound; the positions of 5NSA, Tet, and metal binding residues are within coordinate error (**Figure 6.10a**), and there are minor variations throughout the peptide chain each monomer in the asymmetric units. The coordination geometry for both metals is distorted square pyramidal (**Figure 6.10a**), where the hydroxyl substituent of either 5NSA or Tet occupies the fifth coordination site. Thus, the catalytically competent metal center has octahedral geometry where the fifth position is occupied by the amino group of 5NAA and the sixth coordination site is the water molecule to be activated for hydrolysis (**Figure 6.10b**).

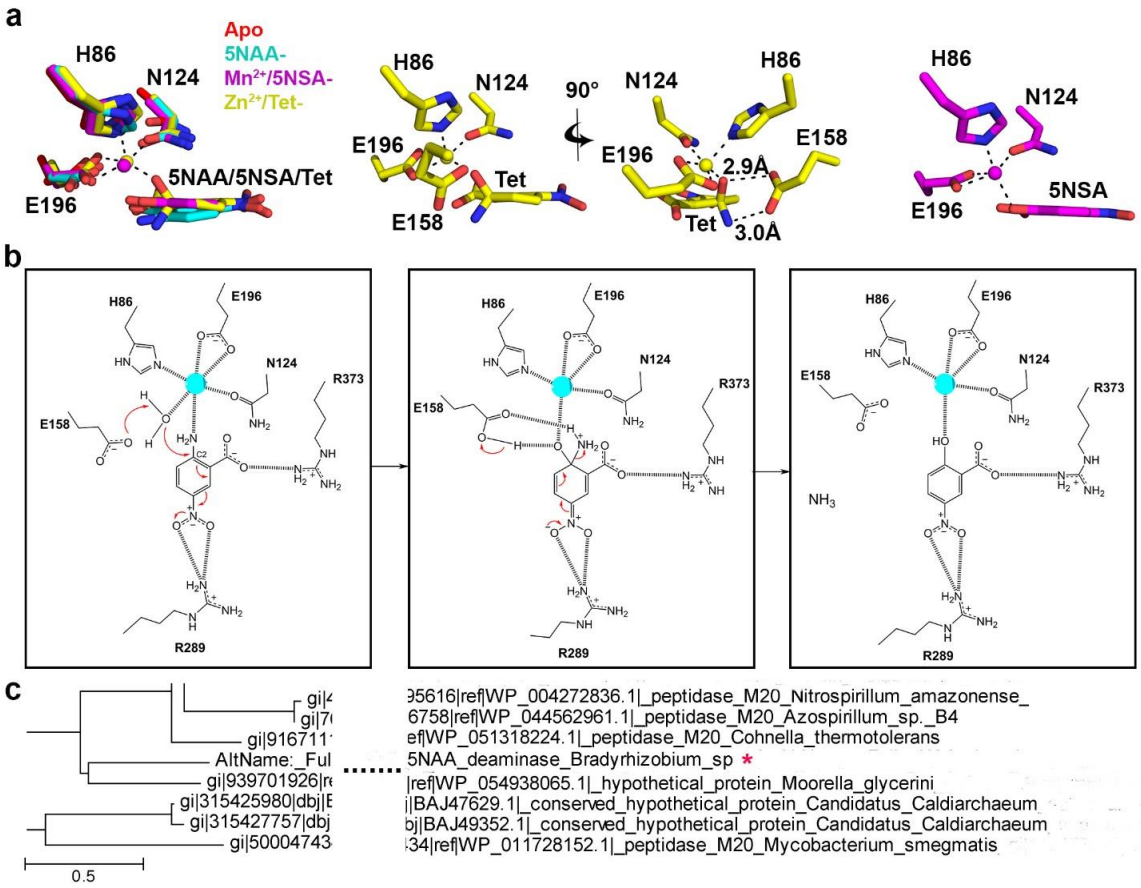


Figure 6.10. Metal coordination, proposed catalytic mechanism and, phylogenetic analysis of 5NAA-A. (**a**) Left: Superposition of metal binding residues of apo 5NAA-A (red), 5NAA-bound (cyan), Mn^{2+}/5NSA-bound (magenta) 5NAA-A and Zn^{2+}/Tet-bound 5NAA-A_{R289A} (yellow). Middle: Coordination environment of Zn^{2+}/Tet-bound 5NAA-A_{R289A} shown in two orientations, highlighting stabilization of the intermediate by Glu_{158}.

Right: Coordination environment for Mn^{2+}/5NSA-bound 5NAA-A. (**b**) Proposed nucleophilic aromatic substitution mechanism for hydrolysis of 5NAA to 5NSA by 5NAA-A. (**c**) Phylogenetic tree (based on protein sequence) zoomed into the region closest to 5NAA-A (*). Full phylogenetic tree is available as a separate file from Supplementary Information.

6.3.4. Proposed chemical mechanism and substrate specificity

On the basis of the available structures, we propose that 5NAA-A-mediated hydrolysis of 5NAA to form 5NSA occurs by S_NAR in a sequence of events that parallels, but is distinct from, metal-mediated proteolysis (*184*) (**Figure 6.10b**). A water molecule is first activated by metal coordination by reducing its pK_a, and then deprotonated by the general base Glu_{158}. Nucleophilic attack of 5NAA at the electrophilic C2 carbon can only occur from a single face, forming the Meisenheimer intermediate with (*R*-) absolute configuration. The now-protonated Glu_{158} stabilizes the hydroxyl and amino groups of the tetrahedral C2 carbon (**Figure 6.10a, b**). Aromaticity is restored concomitant with the release of the ammonia leaving group to form 5NSA. This mechanism is distinct from unrelated DNA deaminases (*161*). Both enzymes activate water via a mononuclear metal ion site, but different ligands, coordination environments, and active site residues accommodate the purine/pyrimidine and corresponding reactive intermediates (*185*). The metal dependence of 5NAA-A renders it also distinct from glutathione-S-transferase and 4-chrlorobenzoyl-CoA dehalogenase, the two other enzymes proposed to undergo S_NAR (*186*).

Interestingly, while the electrons in Tet are expected to be primarily delocalized onto the strongly electron withdrawing *para* nitro group, the *ortho* carboxy substituent of 5NAA is also important for catalysis. 5NAA-A was not reactive toward *m*-nitrobenzoic acid, 4-nitroaniline (4NA), anthranilic acid, aniline, 2-bromo-5-nitrobenzoic acid

(BNBA), 2,4-dinitroaniline, aminoterephthalic acid, or 4-amino-3-nitrobenzoic acid (ANB) (not shown). No binding was detected by ITC for analogs 4NA or ANB (**Figure 6.9b, c**), and neither 4NA, ANB, nor BNBA is a competitive inhibitor (**Figure 6.9e**). The result for ANB is particularly surprising because it has both functional groups that in 5NAA interact with an arginine. Taken together, 5NAA-A exhibits a high level of substrate specificity.

6.3.5. Phylogenetic analysis

Phylogenetic analysis of the 5NAA-A protein sequence expands our appreciation for the breadth of non-peptidase M20 homologs lurking among this protease family far beyond deacetylases and desuccinylases (*50*) (**Figure 6.10c**). Though numerous sequences are annotated as M20 peptidase family members, the number of metal binding sites in the active site and biological substrates are generally not known. Manual inspection of sequence alignments of closely related species reveals just one other sequence, *Moorella glycerini* hypothetical protein WP_054938065.1, with the signature residues for both mononuclear metal ion binding and binding of a nitroaromatic substrate (**Figure 6.10c** and **Figure 6.11**). The biological relevance of degradation of 5NAA-like compounds in this thermophilic, anaerobic, endospore-forming bacterium isolated from a hot spring sediment in Yellowstone National Park (*187*), is a mystery. Review of other selected sequences reveals several additional related enzymes from endospore-forming bacteria, e.g. *Cohnella thermotolerans* peptidase M20 WP_051318224 and *Nitrospirillum amazonense* peptidase M20 WP_004272836.1 (**Figure 6.10c**), which have the same 5NAA-A metal binding residues and may hydrolyze aromatic substrates (Phe in position

of Tyr_{223}) harboring a carboxylate (Arg in position Arg_{373}), but are unlikely to stabilize a nitro group (Gly or Ala in place of Arg_{289}) as in 5NAA-A (**Figure 6.11**).

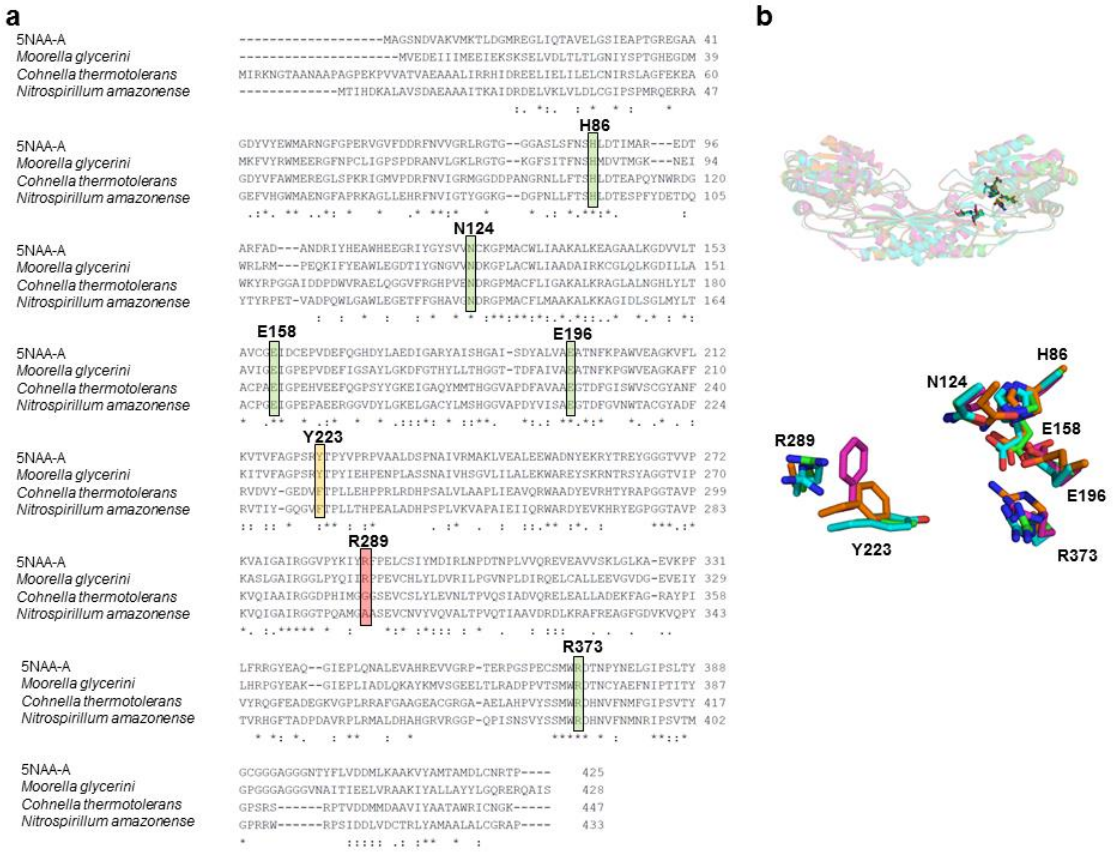


Figure 6.11. Sample of close homologs of 5NAA-A based on phylogenetic analysis. (**a**) Multiple sequence alignment of 5NAA-A with *Moorella glycerini* hypothetical protein WP_054938065.1, *Cohnella thermotolerans* peptidase M20 WP_051318224 and *Nitrospirillum amazonense* peptidase M20 WP_004272836.1. Metal and substrate binding residues of 5NAA-A are highlighted. Conservation and highlight as in Supplementary Fig. 1. (**b**) Overlay of 5NAA-A structure and homology models of homologs generated by using 5NAA-A as a template.

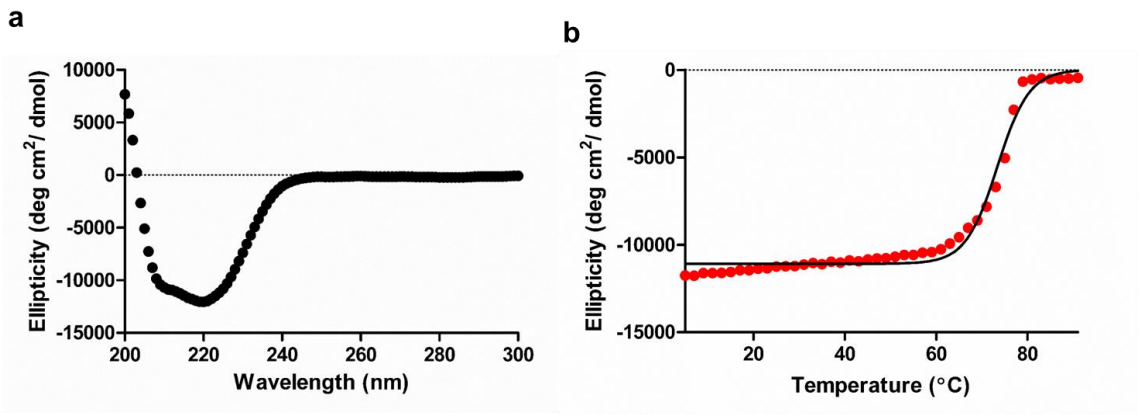


Figure 6.12. Circular Dichroism (CD) analysis of WT 5NAA-A. (**a**) CD spectra (**b**) CD thermal melt with a calculated melting temperature of 74.3°C.

6.4. Discussion

5NAA-A, a notably stable protein (melting temperature of 74 °C, **Figure 6.12**), is well designed for the key first step in biodegradation of the toxic nitroaromatic substrate. 5NAA-A must bind 5NAA before cellular nucleophiles like glutathione, found at milimolar concentrations in the cell (*188*), have an opportunity to form covalent adducts. Although there are no visible channels to the metal site in the 5NAA-bound structure, metal loading *in vivo* must be facile because the enzyme is active in clarified cell lysates of *E. coli* (above) and *Bradyrhizobium sp. JS329* (*159*). No metallochaperone protein sequence is present within the 5NAA-A containing operon (*159*) and the total cellular Zn and Mn levels in *Bradyrhizobium sp. JS329* are in the micromolar range, similar to levels found in *E. coli* (**Figure 6.8c**) (*189*). The ability of 5NAA-A to function with multiple metals suggests that it might rely on available pools of buffered Mn^{2+} or Zn^{2+} (*190*), or those of less prevalent transition metal ions, and is suitably flexible in metal ion utilization so as not to be affected by fluctuations in any one particular metal.

Finally, 5NAA-A expands our knowledge of the largely-uncharacterized metalloproteome (*191*). 5NAA-A would not have been identified as a metal-dependent enzyme by proteome-scale studies, which follow metal content during purification. Whereas bioinformatics correctly predicted 5NAA-A to be a metal dependent enzyme, our findings reinforce the challenges associated with functional annotation of sequences from large families in the absence of known biological substrates (*192*). We identified additional M20 non-peptidase homologs that likely harbor labile mononuclear metal binding sites and hydrolyze other aromatic substrates of potential environmental or synthetic interest. It is not possible to identify substrates from the sequence or structure, but it appears likely that many additional rare and unusual enzymes are involved in catabolic pathways for the great diversity of natural organic compounds in our ecosystem (*54*).

6.5. Future directions

Our cumulative results suggest that a substrate-metal complex likely binds to 5NAA-A. There is strong precedence for microbes to employ a detoxification strategy in which manganese ions form complexes with small organic molecules (*193*), and a similar tactic might be in use by *Bradyrhizobium sp. JS329*.

As one of the referees pointed out in their review of our manuscript, one way to test the substrate-assisted metal loading is to conduct competition experiments using a fluorescence chelator. I have showed that 5NAA can compete out calcein and indo-1 for Zn^{2+}. Reported K_D for Zn^{2+}-indo-1 is 0.16 nM (*194*). This means that 5NAA has a very high affinity for Zn^{2+} with K_D being less than 0.16 nM. It is important to keep in mind

that K_D of 5NAA for Zn^{2+} should be equal or lower than free zinc concentration in the cell for the efficient catalysis.

Future directions for this project include understanding the possible molecular evolution of 5NAA-A from a metallo-peptidase family. Protein engineering efforts can be performed to see if 5NAA-A would revert back and catalyze peptide bond hydrolysis. Unusual metal-coordinating Asn can be mutated to more common Asp to see if enzyme would catalyze the same reaction with same kinetic parameters and if its affinity for metal/substrate would change. Also, active site mutations can be employed to make 5NAA-A accept/catalyze other nitroaromatic substrates that are environmentally and ecologically important.

CHAPTER 7: Perspectives and Future Directions

7.1. Perspectives on membrane protein structure determination

Structures of membrane proteins including Intramembrane Peptidases (IPs) are severely underrepresented. As of May 5th, 2016, there were only 615 unique membrane protein structures deposited in Protein Data Bank (http://www.rcsb.org/) which has 118,587 biological macromolecular structures in total (http://blanco.biomol.uci.edu/mpstruc/). The main reason for this is the challenge associated with crystal growth (Chapter 3).

Because membrane proteins are hard to crystallize with conventional techniques, several new methods to increase the likelihood of obtaining diffraction quality crystals have emerged. One particularly exciting technique using free electron X-ray lasers was recently developed and it introduced a new concept named femtosecond nanocrystallography for membrane protein crystallization (*195, 196*). This method does not need fully-grown crystals, it uses stream of nanocrystals for data collection and structure determination. The method thus overcomes challenges related to rate-limiting steps of crystal growth and radiation sensitivity. This method can also be used for Lipidic Cubic Phases (LCP) grown microcrystals with a specialized LCP injector (*197*). As discussed in Chapter 5, LCP is another recent technique developed for membrane protein crystallization. It uses native-like three-dimensional lipid bilayer system for so called *in meso* crystallization instead of detergent environment (*136*). LCP has gained considerable popularity particularly for GPCRs (*198-201*). A third approach which combines conventional crystallization with native-like environment is Lipopeptide Detergents

(LPDs) (*202*). LPDs are lipid-mimicking detergents consisting of an amphipathic α-helix supported by alkyl chains at both ends of the α-helix (*203*).

In recent years, cryo-EM became extremely powerful in membrane protein structure determination. The data resolution improved a lot over the recent years from medium to high resolution (*204*). Disadvantages of membrane protein crystallization such as need for high amount of pure protein and poor diffraction quality can be overcome with cryo-EM. There is an increasing number of cryo-EM structures, this technique has been used to solve more than 100 protein structures so far (*205*).

7.2. Perspectives on structural and biophysical studies of intramembrane peptidases

Intramembrane proteolysis is unusual in that hydrolysis of peptide bonds occurs within a hydrophobic lipid bilayer environment, presumably by utilizing bulk water. Although these cleavage events are involved in essential roles in various biological processes, little is known about the biochemical/biophysical/structural properties or details of the catalytic mechanisms of intramembrane proteases (IPs). With the availability of the first SPP structure (*42*), some key insights have been gleaned regarding the overall architecture, but most questions surrounding function remain unanswered (Chapter 1).

Structure-function studies systematically probe structure usually by site-directed-mutagenesis, to compare function with the wild-type version of the target protein. Such differences could be related to enzyme kinetics, substrate binding affinity or kinetics, substrate specificity, and interactions with other proteins. Due to the presence of detergents and/or lipids, biochemical investigations of IPs have been largely limited to

discontinuous assays where catalytic parameters and substrate specificity are mostly absent (*68*). Application of modern biophysical techniques such as Förster resonance energy transfer (FRET) assay (*68*) and surface plasmon resonance (SPR) (*206*) would give additional details, but have not been implemented for the study of IPs.

7.3. A preliminary structure-function study of MCMSPP to probe its substrate gating

7.3.1. Introduction

7.3.1.1. Substrate gating for IPs

A major unanswered question about IPs is which structural elements are important for substrate recognition and processing (Chapter 1). For presenilin, the catalytic subunit of γ-secretase responsible for the generation of amyloid β-peptides of different lengths and pathogenicity in Alzheimer's disease (*33*), studies probing substrate gating have demonstrated that the α-helical region of its hydrophilic loop 1 and its C-terminal transmembrane helices are important for the cleavage: a conformational change occurs around these regions upon substrate binding (*207-209*). Conformational changes were also observed in TM2, TM6 (*210*) and TM9 (*211*) upon binding of substrate or transition state analog inhibitor. A lateral gate could allow for substrate access (*212, 213*) but roles of responsible residues and detailed structural mechanism are still unknown.

Presenilin and Signal Peptide Peptidase (SPP) are two members of intramembrane aspartyl peptidase family and they share common structures (Chapter 2) (*214*). There is a conserved Pro-Ala-Leu (PAL) motif both in presenilin and SPP in the aforementioned C-terminal region, the functional importance of which derives from several studies. First,

according to one study investigating the effect of transition-state analogs on the conserved catalytic GxGD and PAL motifs, water accessibility of all residues around those motifs decreased significantly in the presence of inhibitor (*215, 216*). Second, the presenilin PAL motif was shown to be in close proximity of the catalytic center via a cross-linking experiment (*217*). Third, while it was initially proposed that PAL motif was important for both stabilization of γ-secretase multiprotein complex and catalytic activity of presenilin (*218*), later studies indicate it is only important for the activity and substrate gating, not complex formation (*219*). To the best of our knowledge, there is only one study which investigates the importance of PAL motif in SPP activity, which shows that mutation of these residues inhibits the activity of SPP and abolishes its binding to a transition state analog, similar to presenilin (*220*).

7.3.1.2. Chapter overview and publications

In this chapter, I present preliminary data for a structure-function study of MCMSPP substrate gating, which, compared to studies of presenilin, is underrepresented in the literature. Studies probing the PAL motif role in substrate gating of presenilin were conducted using discontinuous assays where enzyme and substrate binding kinetics were not investigated. By studying MCMSPP enzyme kinetics and substrate binding kinetics with different mutants, we can differentiate between catalytic impairment due to an incomplete active site versus impaired substrate binding to the enzyme. Notably, MCMSPP harbors the motif "AGL", not "PAL", which is conserved among archaeal SPPs (**Figure 7.1A, B**). We hypothesize that similar to presenilin, the substrate enters the active site via the interface between catalytic motif ("YD")-harboring TM6 and TM9; to do so, it first encounters residues in and surrounding the PAL motif to reach active site

(**Figure 7.1C**). This might be accompanied by a conformational change, as also proposed for presenilin (section 7.3.1.1). Expansion and completion of this study will be carried out by Swe-Htet Naing (a current PhD candidate in Lieberman Lab) and manuscript will be subsequently written and published.

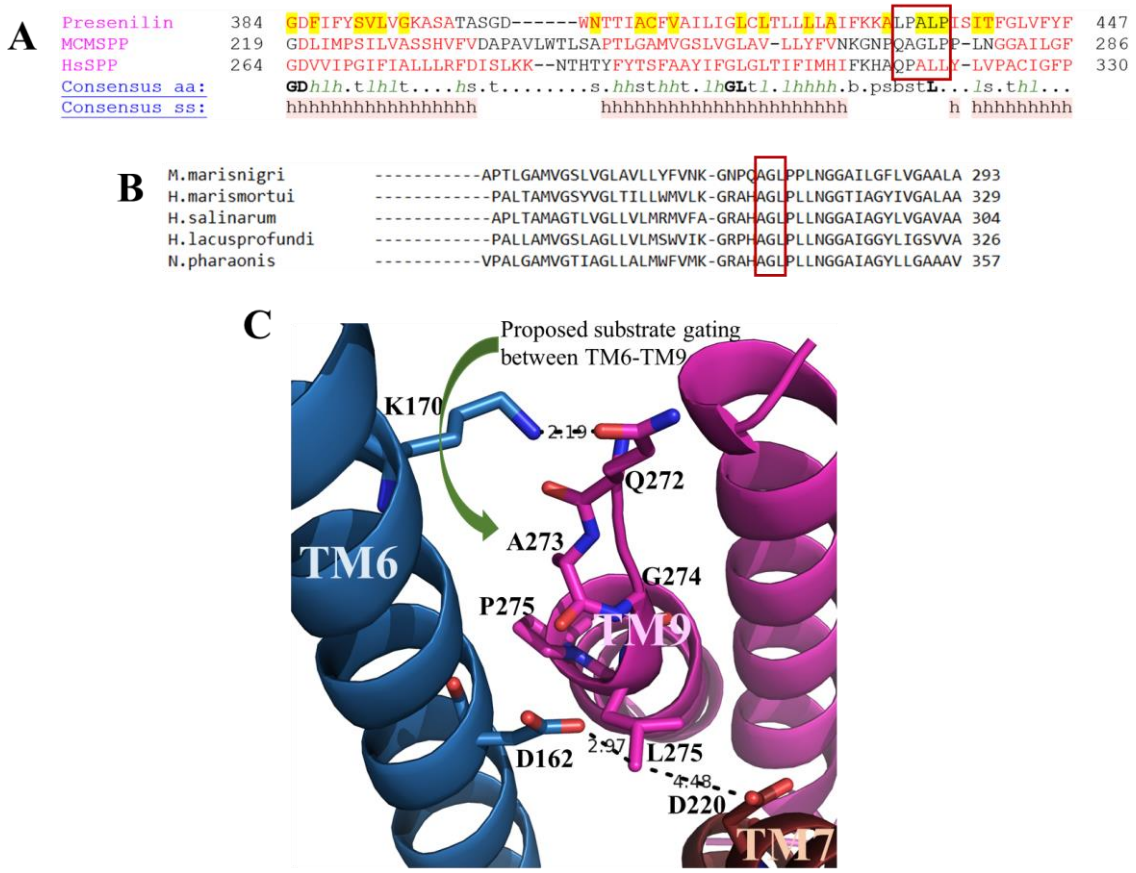


Figure 7.1. Proposed substrate gating mechanism for MCMSPP. (**A**) Multiple sequence alignment of human presenilin, human SPP (HsSPP) and MCMSPP. Conserved PAL motif and a residue before and after PAL motif are boxed in red. Yellow highlights show Familial Alzheimer's Disease (FAD) mutation residues of human presenilin. (**B**) Multiple sequence alignment of archeal SPPs. Conserved AGL motif is boxed in red. (**C**) Structure of catalytic region of MCMSPP (PDB ID: 4Y6K). Catalytic aspartates, residues selected for this structure-function study and corresponding TMs were labeled. Proposed substrate gating between TM6 and TM9 was also labeled.

7.3.2. Methods

7.3.2.1. Generation of MCMSPP variants

Residues expected to affect substrate gating were determined by comparing published structures of MCMSPP and presenilin (*42*, *213*). Selected amino acids were mutated to either Familial Alzhemier's Disease (FAD) mutations extracted from Alzheimer Disease & Frontotemporal Dementia Mutation Database (*221*) or other residues by chemical intuition (**Table 7.1**). Nucleotides were mutated by SDM (Agilent QuickChange II) using full-length WT MCMSPP cloned in the pet22b(+) plasmid as the template (Chapter 2). Correct sequences were confirmed by DNA sequencing (MWG Operon). Primers for all mutants are listed in Table 7.1.

Table 7.1. Primers for MCMSPP mutants.

12	**Mutation**	**Forward (F) and Reverse (R) Primers (5'-3')**
E177A	FAD mutation	F: CACATGATCACGCTGGCCGCGGGCGTCCTCGAGACG R: CGTCTCGAGGACGCCCGCGGCCAGCGTGATCATGTG
P185S	FAD mutation	F: GTCCTCGAGACGAAGGCGTCTATAATGGTCGTGGTTCCG R: CGGAACCACGACCATTATAGACGCCTTCGTCTCGAGGAC
P185L	FAD mutation	F: GTCCTCGAGACGAAGGCGCTCATAATGGTCGTGGTTCCG R: CGGAACCACGACCATTATGAGCGCCTTCGTCTCGAGGAC
Q272A	Polar-to-Alanine	F: GTCAACAAGGGCAACCCCGCGGCGGGTCTCCCCCCC R: GGGGGGGAGACCCGCCGCGGGGTTGCCCTTGTTGAC
A273P/ G274A	MCMSPP-to-Presenilin motif	F: GTCAACAAGGGCAACCCCCAGCCCGCGCTCCCCCCCTTAAACG R: CGTTTAAGGGGGGGAGCGCGGGCTGGGGGTTGCCCTTGTTGAC
G274C	FAD mutation	F: GGGCAACCCCCAGGCGTGCCTCCCCCCCTTAAACG R: CGTTTAAGGGGGGGAGGCACGCCTGGGGGTTGCCC
G274T	FAD mutation	F: GGGCAACCCCCAGGCGACCCTCCCCCCCTTAAACG R: CGTTTAAGGGGGGGAGGGTCGCCTGGGGGTTGCCC
L275F	FAD mutation	F: GGGCAACCCCCAGGCGGGTTTTCCCCCCTTAAACGGCG R: CGCCGTTTAAGGGGGGAAAACCCGCCTGGGGGTTGCCC
P276S	FAD mutation	F: CCCCCAGGCGGGTCTCTCTCCCTTAAACGGCGGGGC R: GCCCCGCCGTTTAAGGGAGAGAGACCCGCCTGGGGG
P276Q	FAD mutation	F: CCCCCAGGCGGGTCTCCAGCCCTTAAACGGCGGGGC R: GCCCCGCCGTTTAAGGGCTGGAGACCCGCCTGGGGG

7.3.2.2. Cell growth, membrane isolation and purification

WT MCMSPP and MCMSPP mutants were expressed, their membranes were isolated and proteins were purified as in Chapters 2.2.2, 2.2.3 and 2.2.4 respectively. In

order to eliminate more impurities (e.g. AcrB), a 5% wash step of imidazole (44 mM) after injection and prior to elution from the Ni^{2+}-affinity column was used for all purifications, as explained in Chapter 2.2.4.

I grew cell paste and isolated membranes for all mutants listed in **Table 7.1**. Because we decided to focus on mutants around "PAL" motif, I did at least one purification and further characterization for all mutants except E177A, P185S and P185L. I had very low protein yield for some of the mutants (G274T, A273P/G274A), so I solubilized more membrane (~1g).

7.3.2.3. Protein stability analysis by circular dichroism (CD) melt

Each purified sample was concentrated using a 50K MWCO Amicon filter to 5 µM. Protein concentration was measured by absorbance at 280 nm using a calculated extinction coefficient. CD spectra and thermal melts were acquired on a Jasco J-810 spectropolarimeter equipped with Neslab RTE 111 circulating water bath and a Jasco PTC-4245/15 temperature control system. CD spectra were taken at room temperature for both gel filtration buffer (blank) and protein sample in gel filtration buffer. CD thermal melt was performed utilizing a 1 °C/min increase in temperature from 5 to 89°C. Both CD spectra and thermal melt (temperature points for every 2°C) were acquired with 15 averaged scans from 300 to 200 nm at a 200 nm/min scan rate, using a 0.1-cm cuvette.

CD spectra for each mutant were plotted from 300 nm to 200 nm. For each mutant, the temperature versus the normalized observed ellipticity (degrees) was plotted at the minimum wavelength (222 nm for α-helical proteins). Because membrane proteins are very stable in protein-detergent-complex, we could not reach the plateau at higher end of temperature to obtain a good sigmoidal fit. Therefore, melting temperatures were not

reported. Both CD spectra and melting curves were plotted using GraphPad Prism 5 software.

7.3.2.4. Activity assays for enzyme kinetics analysis

Enzyme assays for all purified mutants were conducted as described in Chapter 5.3.4 (*68*). Data were fit using the Michaelis−Menten analysis within GraphPad Prism 5. Initial velocities were calculated by the blank-subtracted fluorescence reading over the first 20 min plotted against corresponding substrate concentrations. The Michaelis−Menten equation is $V = V_{max}[S]/(K_m + [S])$ where V is the initial velocity, V_{max} is the maximum velocity at infinite substrate concentration, [S] is substrate concentration, and K_m is the substrate concentration required to reach half the maximal velocity.

Three independent samples of purified enzyme were tested for G274T and two independent samples of purified enzyme were tested for WT, L275F, Q272A. Only one independent sample of purified enzyme was tested for the rest of the mutants (A273P/G274A, G274C, P276S, P276Q).

7.3.2.5. Surface plasmon resonance (SPR) assay for analysis of substrate binding kinetics and affinity

Kinetic substrate binding assays were performed with a BIAcore T200 (GE Healthcare) instrument using Series S Sensor Chip NTA (GE Healthcare). The instrument was located in Organic Cleanroom at Georgia Tech. Each C-terminal hexahistidine-tagged WT or mutant MCMSPP was immobilized onto the NTA chip via Ni^{+2}/NTA chelation to a level of ~1500 Response Unit (RU) as bait for substrate peptides. Running

buffer supplemented with DDM (10mM HEPES, pH 7.4, 150mM NaCl, 0.1% DDM) was also used for Ni^{2+} solution (0.5mM $NiCl_2$), regeneration solution (350mM EDTA), wash solution (3mM EDTA) and protein/peptide dilutions. Ni^{2+} solution was used to charge the NTA chip, regeneration solution was used to regenerate NTA chip between each run cycles (see "Instruction 22-0607-37 AG" document on GE Healthcare website for further details).

The kinetics wizard within the BIAcore software was used to build the experimental program. The program started with one regeneration cycle (with regeneration buffer) and 2-3 start-up cycles (with running buffer). 50 μg/ml protein sample (ligand) was captured with 120 sec contact time at a flow rate of 10 μl/min. The naked Ren390 peptide (analyte, IHPFHLVIHT, lacking fluorophore and quencher; molecular weight of 1369.7 g/mol) was injected from 2 to 40 μM (0, 2, 5, 10, 20, 30, 40 μM) at a flow rate of 10 μl/min with 90 sec of association time and 180 sec of dissociation time. The 2 μM concentration was eliminated from the analysis and/or experiment because it was very close to the baseline and it did not significantly improve the data analysis. The 20 μM sample was injected twice to check the stability of analyte injection (RU responses for those two 20 μM injections should overlay well to continue with the data analysis). Regeneration was performed after each run with a single 180 sec contact time at a flow rate of 30 μl/min and a stabilization period of 60 sec. All experiments were performed at 25 °C.

Before data fitting, spikes at the start and end of injections were removed (see "Biacore Assay Handbook" for further details). The association rate constant (k_{on}), dissociation rate constant (k_{off}) and equilibrium dissociation constant (K_D; $K_D = k_{off}/k_{on}$)

were calculated assuming a Langmuir 1:1 binding model and local R_{max} with the BIAevaluation software. Chi^2 (Chi^2 (RU^2) / R_{max}) and U-value was used to assess the quality of the fit. For reliable data, Chi^2 (accuracy of fitting) should be lower than 0.5 and U-value (statistical significance of parameters) should be lower than 25. BIAcore evaluation software and GraphPad Prism 5 were used for graphical representation of raw and fitted data.

7.3.2.6. In-gel cleavage assay for mass spectrometry (MS) analysis of cleavage site

Freshly prepared enzyme (10-16 μM) was mixed with 4 μM fusion protein MRS substrate (MBP-IHPFHLVIHT-SUMO-His_6) (*68*) in 10 mM sodium phosphate, pH 7.2, 150 mM NaCl (PBS), 0.05% DDM (w/v). The reaction was continued for 48 hours at 37 °C. The reaction was stopped by the addition of ice-cold acetone (6.5 × sample volume) and the mixture was incubated at -20 °C for at least 2 hours. Samples were centrifuged at 8000 x g at 4 °C for 5 min followed by decanting the acetone and air-drying at RT for 30 min. Reducing laemmli sample buffer was added to the precipitated materials, which were then boiled in a 95 °C water bath for 15 min, and loaded onto a 12% (w/v) polyacrylamide gel for SDS-PAGE analysis and coomassie staining. For analysis of cleavage site, LC-MS/MS analysis will be performed for each cleaved band as explained previously. (*68*).

7.3.3. Results

7.3.3.1. Purification and protein yield

MCMSPP (WT and 7 mutants) were purified at least once (**Table 7.2**). Same membrane amount (~0.4g) was used for most of the mutant purifications until low protein

yield was the issue with two of the mutants, G274T and A273P/G274A (**Table 7.2**). Membrane amount was doubled for some purifications to get higher protein yields (**Figure 7.2**). MCMSPP variants eluted from HiPrep 16/60 Sephacryl S-300 column (S300) at ~65-70 ml and their peaks appeared Gaussian, except G274C and G274T where there was an obvious shoulder around 60 ml (**Figure 7.2**). According to SDS-PAGE, the shoulder contained mostly the variant of interest plus some higher molecular weight contaminants (**Figure 7.3**)

Table 7.2. Number of purifications and protein yield for WT and 7 mutants.

MCMSPP variant	Number of purifications	Protein yield (mg/g of membrane)
WT	2	1.05
L275F	2	2.55
Q272A	2	1.17
G274T	3	0.50
G274C	1	0.74
A273P/G274A	1	0.67
P276S	1	1.38
P276Q	1	2.04

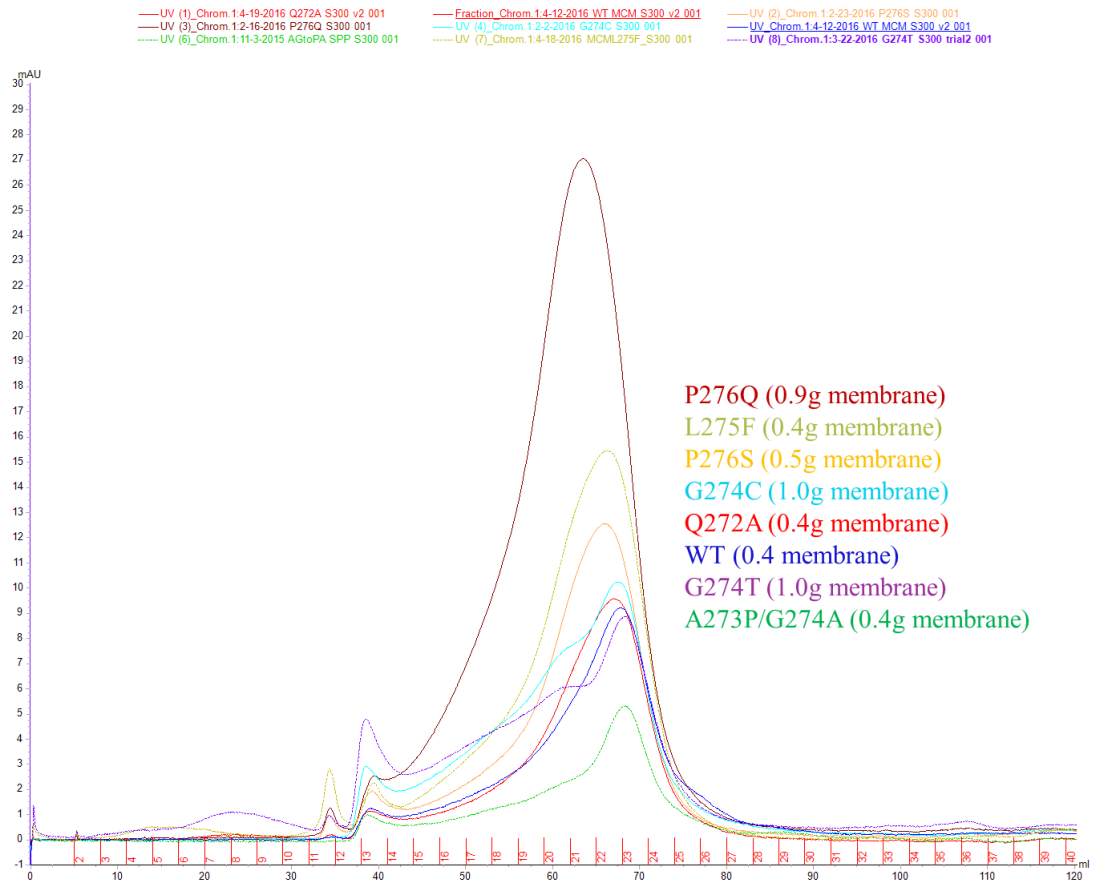


Figure 7.2. Overlayed S300 chromatograms for WT and 7 mutants. Different amounts of membrane were used as stated in the legend. Y-axis is absorbance reading at 280 nm in mAU from AKTA FPLC instrument and x-axis is the elution volume in ml.

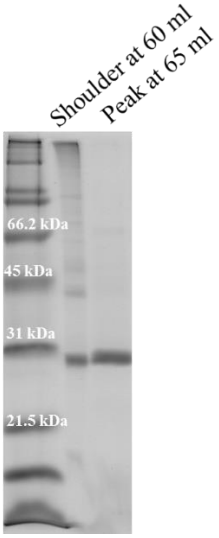


Figure 7.3. SDS-PAGE analysis of shoulder (60 ml) and elution peak (70 ml) of G274T mutant.

According to SDS-PAGE analysis, WT and 7 mutants were pure enough for downstream experiments (**Figure 7.4**). Notably, G274C and A273P/G274A showed a second band between 45 and 66.2 kDa molecular marker bands. Because SPP always runs right below 31 kDa, we surmised this might be a dimer. Mass spectrometry analysis (Systems Mass Spectrometry Core at EBB) identified both bands as corresponding G274C or A273P/G274A (data not shown), indicating the higher band is a dimer. No reducing agent (e.g. β-mercaptoethanol) was present in laemmli sample buffer used for SDS-PAGE because WT MCMSPP does not have any cysteines for possible disulfide bond formation. Thus, while the presence of dimer for G274C could be due to lack of reducing agent, the A273P/G274A dimer is apparently SDS-resistant.

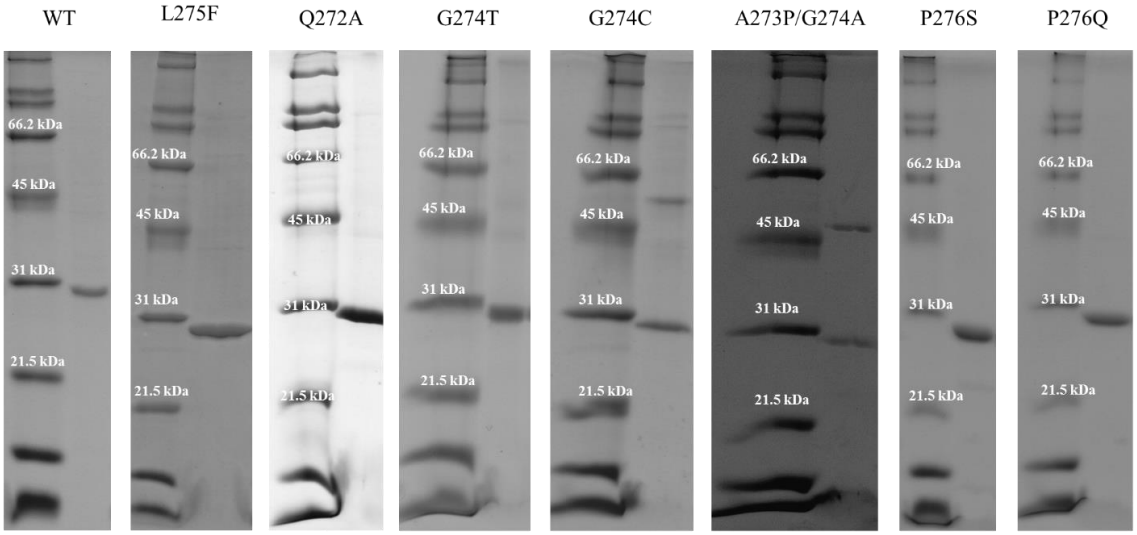


Figure 7.4. SDS-PAGE analysis of final protein samples (after S300 GF) used for further experiments.

7.3.3.2. CD spectra and thermal melt

After each purification, CD spectra and melt were run to make sure proteins were folded and stable enough for downstream experiments. According to their CD spectra conducted at 5μM, all mutants had similar α-helix content as WT except G274T and

A273P/G274A which showed less (**Figure 7.5A**). The only difference between G274T, A273P/G274A and the rest of mutants were their lower protein yield (**Table 7.2**). Because of their lower yield, more concentration steps were required to achieve the same protein concentration, which might affect the final concentration of detergent in the sample (Chapter 2). To see whether protein yield/detergent amount was the issue behind this CD spectra, I performed another purification and CD spectra for G274T using higher starting membrane quantities (**Figure 7.5B**, G274T_1 used 0.4 g membrane, G274T_2 used 1 g membrane). The α-helix content of G274T from second purification was similar to that of WT. This might mean that too much detergent affects the structure of the protein in protein-detergent complex.

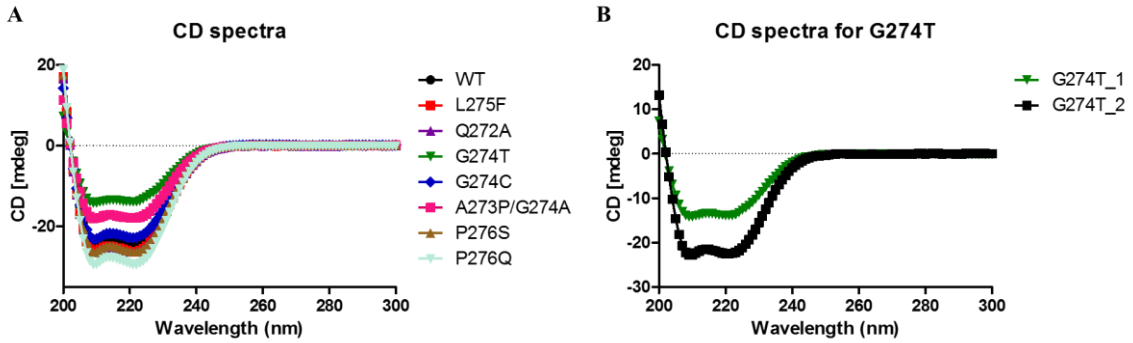


Figure 7.5. CD spectra of (**A**) WT and 7 mutants (**B**) G274T with two batches of purification. G274T_1 used 0.4 g membrane with more filtration steps for final sample and G274T_2 used 1 g membrane with similar number of filtration steps to that of WT.

CD thermal melts were conducted for WT and 7 mutants from 5 to 89 °C and changes in α-helix content were monitored at 222 nm (**Figure 7.6**). The CD signal was normalized because some of mutants had lower apparent CD signal (see above). The data do not conform to a sigmoidal fit characteristic of a two-state unfolding protein, which

could be due to the high stability of protein-detergent complexes (Chapter 2), but melts overlay with one another and WT, suggestive of similar stability.

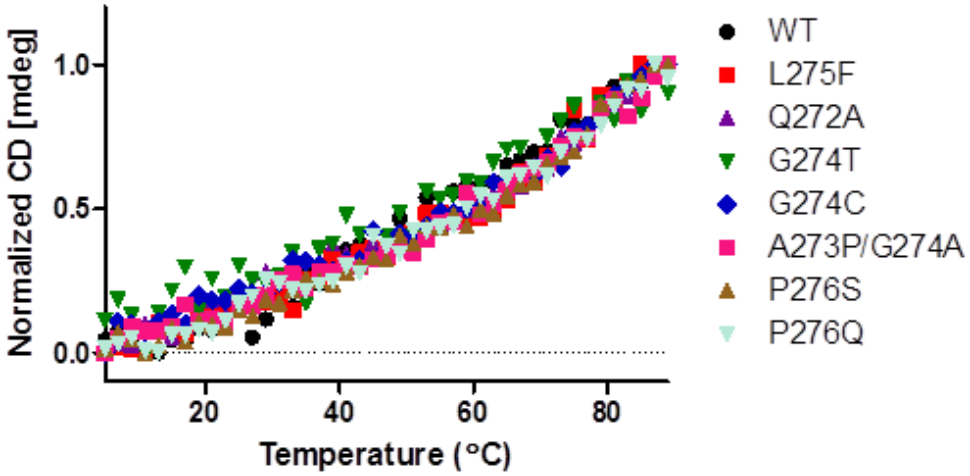


Figure 7.6. Normalized CD thermal melts of WT and 7 mutants at 222 nm.

7.3.3.3. Enzymatic assays and Michaelis-Menten kinetics

Activity assays of WT and all 7 mutants were performed in triplicates (analytical replicates) immediately after purification (**Figure 7.7**). Assays of WT, L275F, Q272A and G274T were repeated with second batches of purified protein because they were selected as a high priority for further testing according to their preliminary data (see Discussion).

All mutations altered the catalytic activity of MCMSPP toward the Ren390FRET substrate. They were all slower than WT except A273P/G274A, which had slightly higher k_{cat} but notably higher K_m (**Table 7.3**). When specificities/efficiencies (k_{cat}/K_m) are compared, all mutants were less efficient than WT.

The first (low yielding) purification of G274T mysteriously yielded a 3-fold higher k_{cat} (data not shown). For this sample, samples were extensively concentrated (60x) to get the desired concentration of protein. I repeated its purification two more

times with more membrane to make it comparable to rest of the mutants (30x filtration). For both purifications, its kinetics were slower than WT (**Table 7.3**, **Figure 7.7**) and parameters were consistent between those two purifications. The reason for the aberrant behavior of the first purification is unknown; it might be due to the altered sample handling, results were subsequently omitted from this study. For the final purification attempt of G274T, I further tried to replicate the first purification condition by setting aside some protein after S300 gel filtration and applying 60x concentration as in the first sample. Kinetics were indeed faster than the remaining G274T but still slower than WT (data not shown). Although it is not the exact replication of first purification, we infer that kinetics are sensitive to concentration steps and likely related to the detergent concentration. In sum, in order to fairly compare kinetics data, mutants should be purified and concentrated as closely as possible to one another.

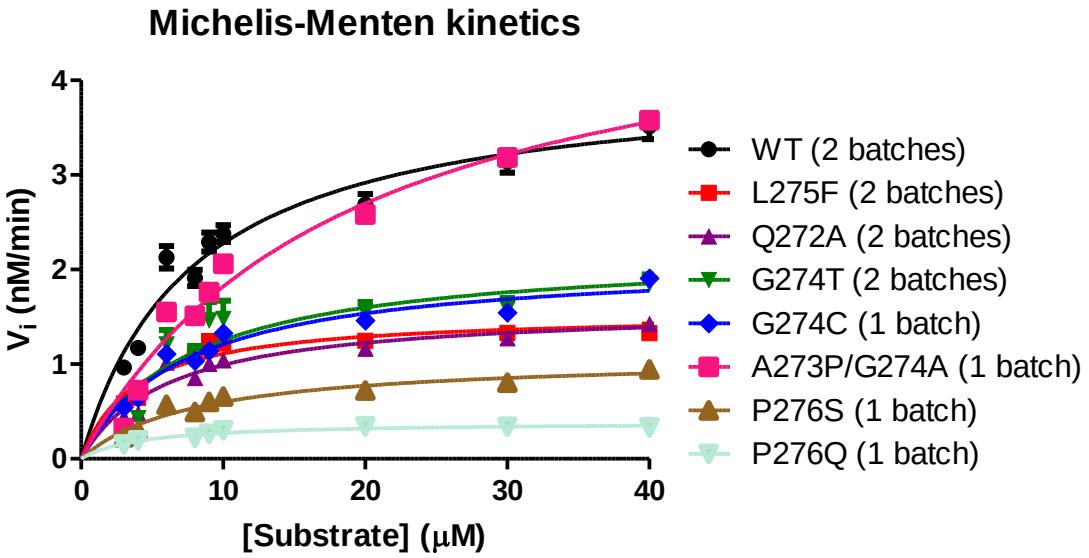


Figure 7.7. Michaelis-Menten analysis of WT and 7 mutants. Some traces represent averages of two independent batches of purified proteins, as stated in the legend.

Table 7.3. Michaelis-Menten parameters of WT and 7 mutants.

MCMSPP variants	V_{max} (nM/min)	k_{cat} (min^{-1}) $x10^{-3}$	K_m (μM)	k_{cat}/K_m $x10^{-3}$
WT	4.06 ± 0.14	8.12 ± 0.28	7.84 ± 0.74	1.04 ± 0.10
WT at 25 °C	0.98 ± 0.14	1.96 ± 0.28	26.59 ± 7.11	0.07 ± 0.02
L275F	1.54 ± 0.05	3.08 ± 0.10	3.85 ± 0.48	0.80 ± 0.10
Q272A	1.60 ± 0.05	3.21 ± 0.11	6.18 ± 0.61	0.52 ± 0.05
G274T	2.20 ± 0.18	4.41 ± 0.35	7.59 ± 1.65	0.58 ± 0.13
G274C	2.10 ± 0.08	4.21 ± 0.16	7.38 ± 0.77	0.57 ± 0.06
A273P/G274A	5.22 ± 0.21	10.44 ± 0.43	18.67 ± 1.54	0.56 ± 0.05
P276S	1.08 ± 0.04	2.17 ± 0.09	8.09 ± 0.88	0.27 ± 0.03
P276Q	0.39 ± 0.02	0.78 ± 0.04	4.27 ± 0.87	0.18 ± 0.04

We tested the activity of WT at 25 °C because all SPR experiments were done at the same temperature (section 7.3.2.5) and we wanted to ensure that cleavage was not occurring during the association time span of the SPR experiment (90 s). As expected, it was 4-fold slower and much less efficient than at 37 °C (**Table 7.3**, **Figure 7.8**), justifying our use of WT MCMSPP in SPR experiments.

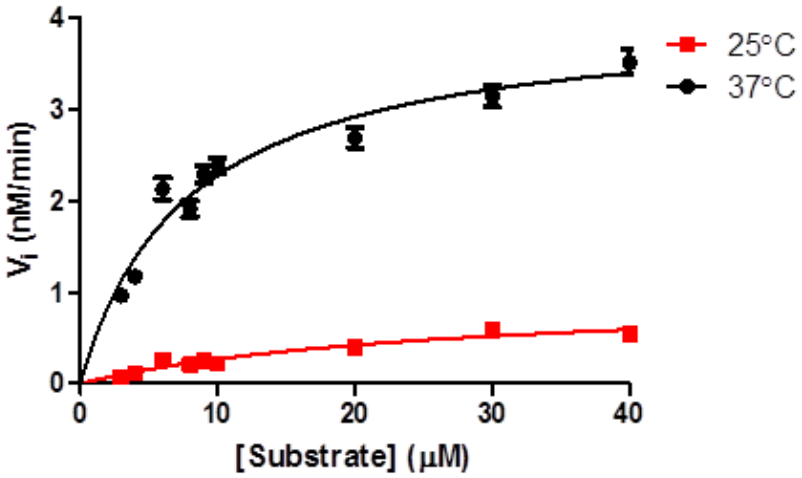


Figure 7.8. Michaelis-Menten analysis of WT at 37 °C and 25 °C.

7.3.3.4. SPR analysis

Because Michaelis-Menten analysis only gives information on reaction kinetics, an SPR assay was developed to determine binding kinetics and affinity. A Ni^{2+}-NTA chip was used to immobilize MCMSPP through its C-terminus His_6 tag and the substrate

corresponding to the peptide sequence of Ren390FRET (without fluorophore/quencher) was tested for binding. All of the assays were run at 25 °C and association time was optimized to 90 seconds. Based on our estimation from reaction kinetics at 25 °C (**Figure 7.8**), during only 90 seconds of association time, cleavage would be insignificant and it would not affect the binding kinetics of the peptide to our protein.

Many parameters were optimized to obtain binding parameters with acceptable statistics. Using WT MCMSPP, parameters tested included: (i) DDM concentration in running buffer (0.05% and 0.1% were tested, 0.1% gave better results), (ii) presence of DDM in Ni^{2+}, regeneration and wash buffers (0.1% DDM is required in all buffers), (iii) regeneration buffer content (350 mM EDTA was better than 500 mM imidazole), (iv) concentration of ligand MCMSPP (25, 50, 100 µg/ml were tried, 50 µg/ml gave the best result with immobilization signal of 1000-1500 RU), (v) concentration range of analyte peptide substrate (120-0 µM and 40-0 µM were tried, 40-0 µM gave better data fit), (vii) association, dissociation times and flow rates (optimal values stated in methods). Subsequently, all mutants were run with the same program and variables.

Figure 7.9 shows the representative SPR sensorgrams of WT and tested mutants; G274T and A273P/G274A could not be tested with SPR due to insufficient yield of purified enzyme. Binding of substrate to tested mutants was measured with only one batch of protein in duplicates; WT MCMSPP was tested with two independent batches of protein with the same binding results (**Table 7.4**). As judged by Chi^2 values from data fitting, only binding values for WT, L275F and Q272A are acceptable. For G274C, P276Q and P276S, reliable fit values are probably restricted by the detection limit of the instrument for k_{on} ("Quality control" tab in Biacore evaluation software, also see

"Biacore T200" document for further details). Binding kinetics rates and K_D values for substrate binding to L275F were similar to WT while Q272A showed significantly higher dissociation rate and lower binding affinity (**Table 7.4**, **Figure 7.10**). However, both of them exhibited slower enzyme kinetics (**Figure 7.7**).

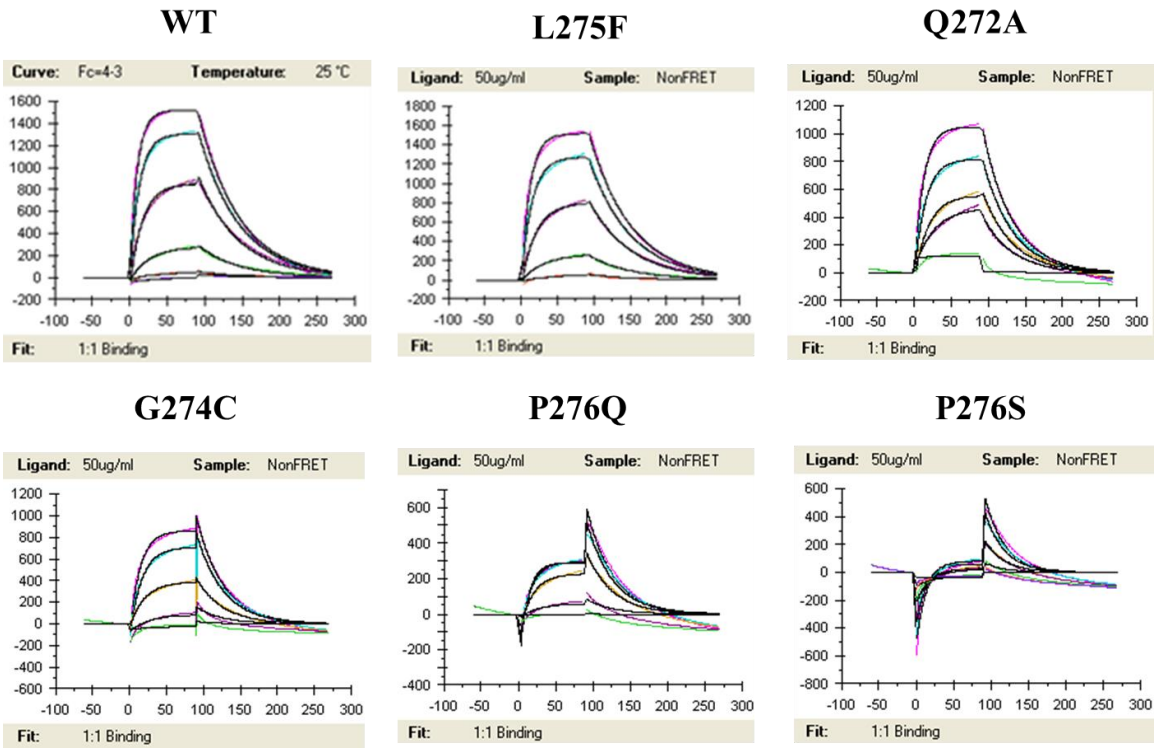


Figure 7.9. Representative SPR sensorgrams of WT and tested mutants. Raw data (black) and fitted data (colored) were overlayed. Response unit (RU) is on y-axis and time in seconds is on x-axis.

Table 7.4. SPR binding kinetics and affinity parameters calculated from data fitting.

Variant	Number of SPR assays	k_{on} (1/Ms)	k_{off} (1/s) x10^{-5}	K_D (µM)	Chi^2	U-value
WT	2 duplicates	2113.8 ± 6.3	1683.8 ± 1.7	7.99 ± 0.02	0.25	0.49
L275F	1 duplicate	1664.0 ± 4.2	1751.0 ± 1.5	10.53 ± 0.03	0.13	0.65
Q272A	1 duplicate	1357.5 ± 12.4	2549.0 ± 7.4	18.81 ± 0.18	0.57	1.65
G274C*	1 duplicate	1097.0 ± 18.4	2875.0 ± 22.4	26.30 ± 0.50	0.91	2.78
P276Q*	1 duplicate	1440.0 ± 30.0	2981.0 ± 21.0	20.79 ± 0.46	1.65	4.08
P276S*	Single	972.8 ± 36.0	3853.0 ± 39.0	39.61 ± 1.52	2.01	6.90

* Parameters should be considered with caution due to their high Chi^2 values (>0.5)

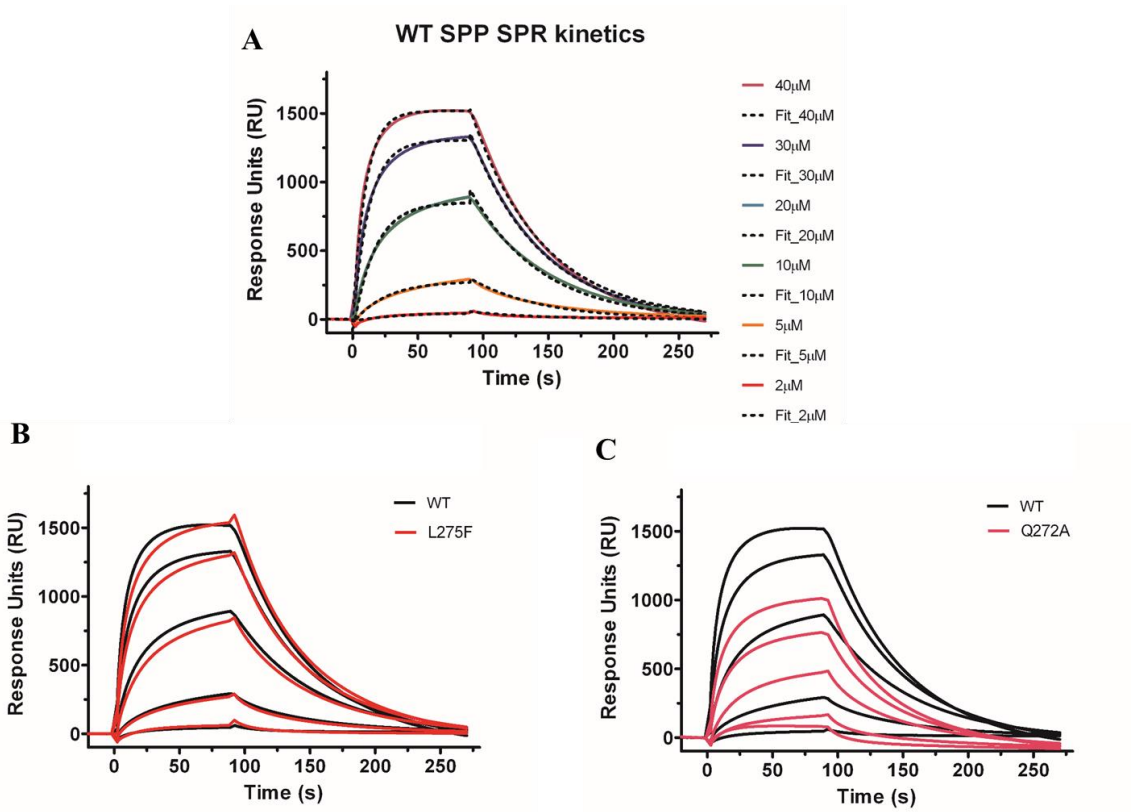


Figure 7.10. SPR sensorgrams of (**A**) WT with raw and fitted data (**B**) WT and L275F overlayed (**C**) WT and Q272A overlayed. (**B**) and (**C**) were plotted with corresponding raw data.

7.3.3.5. Cleavage site

In-gel cleavage assay was performed for the most promising mutants (L275F, Q272A and G274T) to compare the cleavage site preferences. This high resolution

proteomics mass spectrometry experiment is ongoing and conducted in collaboration with the Torres Lab (School of Biology).

7.3.4. Discussion and future directions

This study aims to differentiate residues in MCMSPP that are involved in substrate gating from those directly involved in catalysis. Mutations generated within and adjacent to the conserved so-called PAL motif (**Figure 7.1**, "AGL" in archaeal SPP) were motivated in part by the presence of Alzheimer-associated FAD mutations in this region within presenilin. Included here are preliminary results for seven of the mutants (Q272A, A273P/G274A, G274C, G274T, L275F, P276S, and P276Q). Additional mutants were prepared (E177A, P185S, and P185L) for further study: (a) E177A because it is a conserved charged residue on loop between TM6-TM7 and a FAD mutation (b) P185 because it is in a predicted β-sheet (Chapter 5) and has FAD mutations (P185S, P185L are both documented at this position). Indeed, the function of this conserved predicted β-sheet on TM6-TM7 loop remains obscure as electron density for this feature is absent in the published inactive MCMSPP structure (*42*).

To our knowledge, this is the first SPR analysis of substrate binding to any IP. SPR gave reliable data for WT MCMSPP: K_D values obtained by SPR (7.99 ± 0.02 μM) were in line with our published K_m (7.84 ± 0.74 μM) from the enzyme activity assay and fluorescence polarization binding experiments using the fluorescent substrate (6.5 ± 2.2 μM) (*68*).

Because this study was started at the late stages of thesis work, only preliminary data is presented, of which L275F, Q272A have emerged as the most intriguing. The L275F variant is less efficient and 2-fold slower than WT but surprisingly, its binding

kinetics and affinity remain unchanged. These results demonstrate that L275 is specifically important for the fidelity of enzyme catalysis, not for substrate gating. This effect can be rationalized by the location of L275 between the two catalytic aspartates (**Figure 7.1C**) (*42*). The bulkier FAD mutation L275F retains hydrophobicity but the larger side chain could shift aspartates to a position too far for efficient catalysis. Q272A is 2-fold slower and less efficient than WT, but the SPR result reveals significantly lower k_{on} (slower association) and higher k_{off} (faster dissociation) (**Table 7.4**). Q272, located just N-terminal to the PAL motif on TM9 near the membrane interface, may form a hydrogen bonding interaction with K170 on TM6 (**Figure 7.1C**) to properly orient the two helices for proper substrate binding, which is disrupted in Q272A.

More mutants will be tested by Swe-Htet Naing to investigate the importance of conserved "GXGD" and "YD" motifs which locate two catalytic aspartates. We anticipate that application of the suite of biochemical, biophysical, and bioanalytical methods to the study the structural elements that confer enzymatic function of MCMSPP will continue to provide novel insight into this fascinating enzyme.

REFERENCES

[1] E. Deu, M. Verdoes, M. Bogyo, New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol* **19**, 9-16 (2012).

[2] J. W. Tobias, A. Varshavsky, Cloning and Functional-Analysis of the Ubiquitin-Specific Protease Gene Ubp1 of Saccharomyces-Cerevisiae. *Journal of Biological Chemistry* **266**, 12021-12028 (1991).

[3] H. S. Speno *et al.*, Site-directed mutagenesis of predicted active site residues in glutamate carboxypeptidase II. *Mol Pharmacol* **55**, 179-185 (1999).

[4] M. B. West *et al.*, Novel insights into eukaryotic gamma-glutamyltranspeptidase 1 from the crystal structure of the glutamate-bound human enzyme. *Journal of Biological Chemistry* **289**, 11569-11569 (2014).

[5] C. J. Scott, C. C. Taggart, Biologic protease inhibitors as novel therapeutic agents. *Biochimie* **92**, 1681-1688 (2010).

[6] D. Gabriel, M. F. Zuluaga, H. van den Bergh, R. Gurny, N. Lange, It is All About Proteases: From Drug Delivery to In Vivo Imaging and Photomedicine. *Curr Med Chem* **18**, 1785-1805 (2011).

[7] B. Turk, Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5**, 785-799 (2006).

[8] M. Drag, G. S. Salvesen, Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9**, 690-701 (2010).

[9] Q. Li, L. Yi, P. Marek, B. L. Iverson, Commercial proteases: present and future. *FEBS Lett* **587**, 1155-1163 (2013).

[10] C. Lopez-Otin, C. M. Overall, Protease degradomics: A new challenge for proteomics. *Nat Rev Mol Cell Bio* **3**, 509-519 (2002).

[11] P. M. Quiros, T. Langer, C. Lopez-Otin, New roles for mitochondrial proteases in health, ageing and disease. *Nat Rev Mol Cell Biol* **16**, 345-359 (2015).

[12] B. Turk, D. Turk, V. Turk, Protease signalling: the cutting edge. *EMBO J* **31**, 1630-1643 (2012).

[13] C. Lopez-Otin, J. S. Bond, Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **283**, 30433-30437 (2008).

[14] K. Oda, New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. *J Biochem-Tokyo* **151**, 13-25 (2012).

[15] L. Polgar, The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences* **62**, 2161-2172 (2005).

[16] S. Rowsell *et al.*, Crystal structure of carboxypeptidase G(2), a bacterial enzyme with applications in cancer therapy. *Structure* **5**, 337-347 (1997).

[17] N. D. Rawlings, A. J. Barrett, A. Bateman, Asparagine Peptide Lyases A SEVENTH CATALYTIC TYPE OF PROTEOLYTIC ENZYMES. *Journal of Biological Chemistry* **286**, 38321-38328 (2011).

[18] N. D. Rawlings, A. J. Barrett, Evolutionary Families of Peptidases. *Biochem J* **290**, 205-218 (1993).

[19] G. Kurisu *et al.*, Structure of the zinc endoprotease from Streptomyces caespitosus. *J Biochem-Tokyo* **121**, 304-308 (1997).

[20] G. P. Bertenshaw, M. T. Norcum, J. S. Bond, Structure of homo- and hetero-oligomeric meprin metalloproteases - Dimers, tetramers, and high molecular mass multimers. *Journal of Biological Chemistry* **278**, 2522-2532 (2003).

[21] J. Lowe *et al.*, Crystal-Structure of the 20s Proteasome from the Archaeon T-Acidophilum at 3.4-Angstrom Resolution. *Science* **268**, 533-539 (1995).

[22] C. C. Zhang, S. S. Wu, D. G. Xu, Catalytic Mechanism of Angiotensin-Converting Enzyme and Effects of the Chloride Ion. *J Phys Chem B* **117**, 6635-6645 (2013).

[23] C. Betzel *et al.*, Structure of the Complex of Proteinase-K with a Substrate-Analog Hexapeptide Inhibitor at 2.2-Angstrom Resolution. *Journal of Biological Chemistry* **268**, 15854-15858 (1993).

[24] N. J. Agard, J. A. Wells, Methods for the proteomic identification of protease substrates. *Curr Opin Chem Biol* **13**, 503-509 (2009).

[25] J. C. Timmer *et al.*, Structural and kinetic determinants of protease substrates. *Nat Struct Mol Biol* **16**, 1101-1108 (2009).

[26] M. Vizovisek, R. Vidmar, M. Fonovic, B. Turk, Current trends and challenges in proteomic identification of protease substrates. *Biochimie* **122**, 77-87 (2016).

[27] B. E. Turk, L. L. Huang, E. T. Piro, L. C. Cantley, Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nature Biotechnology* **19**, 661-667 (2001).

[28] A. J. O'Donoghue *et al.*, Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods* **9**, 1095-1100 (2012).

[29] H. Neurath, K. A. Walsh, Role of Proteolytic-Enzymes in Biological Regulation. *P Natl Acad Sci USA* **73**, 3825-3832 (1976).

[30] A. Konovalova, L. Sogaard-Andersen, L. Kroos, Regulated proteolysis in bacterial development. *FEMS Microbiol Rev* **38**, 493-522 (2014).

[31] C. M. Overall, C. Lopez-Otin, Strategies for MMP inhibition in cancer: Innovations for the post-trial era. *Nature Reviews Cancer* **2**, 657-672 (2002).

[32] R. Sherrington *et al.*, Cloning of a Gene Bearing Missense Mutations in Early-Onset Familial Alzheimers-Disease. *Nature* **375**, 754-760 (1995).

[33] M. S. Wolfe, Intramembrane-cleaving Proteases. *Journal of Biological Chemistry* **284**, 13969-13973 (2009).

[34] B. Furie, B. C. Furie, Molecular and Cellular Biology of Blood-Coagulation. *New Engl J Med* **326**, 800-806 (1992).

[35] J. Adams *et al.*, Proteasome inhibitors: A novel class of potent and effective antitumor agents. *Cancer Research* **59**, 2615-2622 (1999).

[36] D. Langosch, C. Scharnagl, H. Steiner, M. K. Lemberg, Understanding intramembrane proteolysis: from protein dynamics to reaction kinetics. *Trends Biochem Sci* **40**, 318-327 (2015).

[37] S. W. Dickey, R. P. Baker, S. Cho, S. Urban, Proteolysis inside the Membrane Is a Rate-Governed Reaction Not Driven by Substrate Affinity. *Cell* **155**, 1270-1281 (2013).

[38] F. Kamp *et al.*, Intramembrane Proteolysis of beta-Amyloid Precursor Protein by gamma-Secretase Is an Unusually Slow Process. *Biophys J* **108**, 1229-1237 (2015).

[39] S. F. Lichtenthaler, C. Haass, H. Steiner, Regulated intramembrane proteolysis - lessons from amyloid precursor protein processing. *J Neurochem* **117**, 779-796 (2011).

[40] E. Y. L. Chan, G. A. McQuibban, The mitochondrial rhomboid protease: Its rise from obscurity to the pinnacle of disease-relevant genes. *Bba-Biomembranes* **1828**, 2916-2925 (2013).

[41] I. Manolaridis *et al.*, Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature* **504**, 301-+ (2013).

[42] X. Li *et al.*, Structure of a presenilin family intramembrane aspartate protease. *Nature* **493**, 56-61 (2013).

[43] Y. C. Wang, Y. J. Zhang, Y. Ha, Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179-183 (2006).

[44] P. L. Lu *et al.*, Three-dimensional structure of human gamma-secretase. *Nature* **512**, 166-+ (2014).

[45] N. D. Rawlings, A. J. Barrett, R. Finn, Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **44**, D343-D350 (2016).

[46] J. A. Gerlt, P. C. Babbitt, Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annual Review of Biochemistry* **70**, 209-246 (2001).

[47] N. D. Rawlings, F. R. Morton, The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* **90**, 243-259 (2008).

[48] Y. Qu, J. C. Spain, Molecular and biochemical characterization of the 5-nitroanthranilic acid degradation pathway in Bradyrhizobium sp. strain JS329. *J Bacteriol* **193**, 3057-3063 (2011).

[49] S. Kalyoncu *et al.*, Molecular adaptation of a metalloprotease for the hydrolysis of 5-nitroanthranilic acid. *Submitted*, (2016).

[50] A. Boyen *et al.*, Acetylornithine Deacetylase, Succinyldiaminopimelate Desuccinylase and Carboxypeptidase-G2 Are Evolutionarily Related. *Gene* **116**, 1-6 (1992).

[51] K. S. Makarova, N. V. Grishin, The Zn-peptidase superfamily: Functional convergence after evolutionary divergence. *Journal of Molecular Biology* **292**, 11-17 (1999).

[52] D. S. Shi, X. L. Yu, L. Roth, M. Tuchman, N. M. Allewell, Structure of a novel N-acetyl-L-citrulline deacetylase from Xanthomonas campestris. *Biophysical Chemistry* **126**, 86-93 (2007).

[53] B. P. Nocek, D. M. Gillner, Y. Fan, R. C. Holz, A. Joachimiak, Structural Basis for Catalysis by the Mono- and Dimetalated Forms of the dapE-Encoded N-succinyl-L,L-Diaminopimelic Acid Desuccinylase. *Journal of Molecular Biology* **397**, 617-626 (2010).

[54] J. Davies, Small molecules: The lexicon of biodiversity. *J Biotechnol* **129**, 3-5 (2007).

[55] M. S. Brown, J. Ye, R. B. Rawson, J. L. Goldstein, Regulated intramembrane proteolysis: A control mechanism conserved from bacteria to humans. *Cell* **100**, 391-398 (2000).

[56] S. Urban, M. Freeman, Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr Opin Genet Dev* **12**, 512-518 (2002).

[57] E. Erez, D. Fass, E. Bibi, How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**, 371-378 (2009).

[58] Y. C. Wang, Y. Ha, Open-cap conformation of intramembrane protease GlpG. *P Natl Acad Sci USA* **104**, 2098-2102 (2007).

[59] A. Quigley *et al.*, The Structural Basis of ZMPSTE24-Dependent Laminopathies. *Science* **339**, 1604-1607 (2013).

[60] B. Martoglio, T. E. Golde, Intramembrane-cleaving aspartic proteases and disease: presenilins, signal peptide peptidase and their homologs. *Hum Mol Genet* **12**, R201-R206 (2003).

[61] T. Sato *et al.*, Signal peptide peptidase: biochemical properties and modulation by nonsteroidal antiinflammatory drugs. *Biochemistry* **45**, 8649-8656 (2006).

[62] A. Weihofen, K. Binns, M. K. Lemberg, K. Ashman, B. Martoglio, Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296**, 2215-2218 (2002).

[63] F. El Hage *et al.*, Preprocalcitonin signal peptide generates a cytotoxic T lymphocyte-defined tumor epitope processed by a proteasome-independent pathway. *P Natl Acad Sci USA* **105**, 10119-10124 (2008).

[64] M. K. Lemberg, F. A. Bland, A. Weihofen, V. M. Braud, B. Martoglio, Intramembrane proteolysis of signal peptides: an essential step in the generation of HLA-E epitopes. *J Immunol* **167**, 6441-6446 (2001).

[65] M. Voss, B. Schroder, R. Fluhrer, Mechanism, specificity, and physiology of signal peptide peptidase (SPP) and SPP-like proteases. *Bba-Biomembranes* **1828**, 2828-2839 (2013).

[66] F. Mancia, J. Love, High-throughput expression and purification of membrane proteins. *Journal of Structural Biology* **172**, 85-93 (2010).

[67] J. L. Johnson, S. Kalyoncu, R. L. Lieberman, in *Methods in Molecular Biology*. (Springer, Clifton, N.J., 2015).

[68] S. H. Naing *et al.*, Catalytic Properties of lntramembrane Aspartyl Protease Substrate Hydrolysis Evaluated Using a FRET Peptide Cleavage Assay. *Acs Chemical Biology* **10**, 2166-2174 (2015).

[69] H. B. Shen, K. C. Chou, Signal-3L: A 3-layer approach for predicting signal peptides. *Biochem Bioph Res Co* **363**, 297-303 (2007).

[70] F. van den Ent, J. Lowe, RF cloning: A restriction-free method for inserting target genes into plasmids. *J Biochem Bioph Meth* **67**, 67-74 (2006).

[71] C. Torres-Arancivia *et al.*, Identification of an Archaeal Presenilin-Like Intramembrane Protease. *Plos One* **5**, (2010).

[72] E. R. Geertsma, M. Groeneveld, D. J. Slotboom, B. Poolman, Quality control of overexpressed membrane proteins. *P Natl Acad Sci USA* **105**, 5722-5727 (2008).

[73] P. Raman, V. Cherezov, M. Caffrey, The membrane protein data bank. *Cell Mol Life Sci* **63**, 36-51 (2006).

[74] R. Mancusso, N. K. Karpowich, B. K. Czyzewski, D. N. Wang, Simple screening method for improving membrane protein thermostability. *Methods* **55**, 324-329 (2011).

[75] G. A. Malawski *et al.*, Identifying protein construct variants with increased crystallization propensity - A case study. *Protein Science* **15**, 2718-2728 (2006).

[76] C. Tribet, R. Audebert, J. L. Popot, Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *P Natl Acad Sci USA* **93**, 15047-15050 (1996).

[77] T. C. Kaufmann, A. Engel, H. W. Remigy, A novel method for detergent concentration determination. *Biophys J* **90**, 310-317 (2006).

[78] C. C. Prince, Z. C. Jia, Detergent quantification in membrane protein samples and its application to crystallization experiments. *Amino Acids* **45**, 1293-1302 (2013).

[79] P. Strop, A. T. Brunger, Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci* **14**, 2207-2211 (2005).

[80] T. Kawate, E. Gouaux, Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673-681 (2006).

[81] D. Drew, M. Lerch, E. Kunji, D. J. Slotboom, J. W. de Gier, Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3**, 303-313 (2006).

[82] Y. Sonoda *et al.*, Benchmarking Membrane Protein Detergent Stability for Improving Throughput of High-Resolution X-ray Structures. *Structure* **19**, 17-25 (2011).

[83] W. N. Price *et al.*, Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol* **27**, 51-57 (2009).

[84] Y. Qutub *et al.*, Crystallization of transmembrane proteins in cubo: mechanisms of crystal growth and defect formation. *J Mol Biol* **343**, 1243-1254 (2004).

[85] C. Hunte, H. Michel, Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol* **12**, 503-508 (2002).

[86] R. L. Lieberman, J. A. Culver, K. C. Entzminger, J. C. Pai, J. A. Maynard, Crystallization chaperone strategies for membrane proteins. *Methods* **55**, 293-302 (2011).

[87] C. Hunte, H. Michel, Crystallisation of membrane proteins mediated by antibody fragments. *Curr. Opin. Struct. Biol.* **12**, 503-508 (2002).

[88] S. Koide, Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.* **19**, 449-457 (2009).

[89] S. G. Rasmussen *et al.*, Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383-387 (2007).

[90] L. C. Kovari, C. Momany, M. G. Rossmann, The use of antibody fragments for crystallization and structure determinations. *Structure* **3**, 1291-1293 (1995).

[91] C. Hunte, J. Koepke, C. Lange, T. Rossmanith, H. Michel, Structure at 2.3 angstrom resolution of the cytochrome bc(1) complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Struct Fold Des* **8**, 669-684 (2000).

[92] V. Tereshko *et al.*, Toward chaperone-assisted crystallography: Protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (V(H)H) scaffold. *Protein Science* **17**, 1175-1187 (2008).

[93] K. A. Smith *et al.*, Demystified ... recombinant antibodies. *J Clin Pathol* **57**, 912-917 (2004).

[94] M. Kaufmann *et al.*, Crystal structure of the anti-His tag antibody 3D5 single-chain fragment complexed to its antigen. *J. Mol. Biol.* **318**, 135-147 (2002).

[95] J. C. Pai *et al.*, Conversion of scFv peptide-binding specificity for crystal chaperone development. *Prot. Engin. Des. Sel.* **24**, 419-428 (2011).

[96] S. Kalyoncu *et al.*, Effects of protein engineering and rational mutagenesis on crystal lattice of single chain antibody fragments. *Proteins-Structure Function and Bioinformatics* **82**, 1884-1895 (2014).

[97] J. J. Lavinder, S. B. Hari, B. J. Sullivan, T. J. Magliery, High-Throughput Thermal Scanning: A General, Rapid Dye-Binding Thermal Shift Screen for Protein Engineering. *J Am Chem Soc* **131**, 3794-3795 (2009).

[98] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307-326 (1997).

[99] A. J. Mccoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

[100] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).

[101] P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352-367 (2012).

[102] J. J. Headd *et al.*, Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D* **68**, 381-390 (2012).

[103] L. Urzhumtseva, P. V. Afonine, P. D. Adams, A. Urzhumtsev, Crystallographic model quality at a glance. *Acta Crystallogr D* **65**, 297-300 (2009).

[104] E. N. Brown, S. Ramaswamy, Quality of protein crystal structures. *Acta Crystallogr D* **63**, 941-950 (2007).

[105] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774-797 (2007).

[106] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D* **66**, 486-501 (2010).

[107] P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213-221 (2010).

[108] S. Koide, S. S. Sidhu, The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem. Biol.* **4**, 325-334 (2009).

[109] M. Cieslik, Z. S. Derewenda, The role of entropy and polarity in intermolecular contacts in protein crystals. *Acta Crystallogr. D* **65**, 500-509 (2009).

[110] Z. S. Derewenda, It's all in the crystals. *Acta Crystallogr.* **D67**, 243-248 (2011).

[111] J. Janin, F. Rodier, Protein-protein interaction at crystal contacts. *Proteins* **23**, 580-587 (1995).

[112] O. Carugo, P. Argos, Protein-protein crystal-packing contacts. *Protein Sci.* **6**, 2261-2263 (1997).

[113] X. J. Zhang, J. A. Wozniak, B. W. Matthews, Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme. *J. Mol. Biol.* **250**, 527-552 (1995).

[114] J. Janin, Specific versus non-specific contacts in protein crystals. *Nat Struct Biol* **4**, 973-974 (1997).

[115] M. P. Crosio, J. Janin, M. Jullien, Crystal Packing in 6 Crystal Forms of Pancreatic Ribonuclease. *J. Mol. Biol.* **228**, 243-251 (1992).

[116] Y. Wine, N. Cohen-Hadar, R. Lamed, A. Freeman, F. Frolow, Modification of Protein Crystal Packing by Systematic Mutations of Surface Residues: Implications on Biotemplating and Crystal Porosity. *Biotechnol Bioeng* **104**, 444-457 (2009).

[117] Y. N. Kang, M. Adachi, B. Mikami, S. Utsumi, Change in the crystal packing of soybean beta-amylase mutants substituted at a few surface amino acid residues. *Protein Eng* **16**, 809-817 (2003).

[118] D. R. Cooper *et al.*, Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. D* **63**, 636-645 (2007).

[119] R. P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, A dissection of specific and non-specific protein - Protein interfaces. *J. Mol. Biol.* **336**, 943-955 (2004).

[120] P. G. Vekilov, A. R. Feeling-Taylor, S. T. Yau, D. Petsev, Solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallogr. D* **58**, 1611-1616 (2002).

[121] J. D. Pedelacq *et al.*, Engineering soluble proteins for structural genomics. *Nat Biotechnol* **20**, 927-932 (2002).

[122] Z. S. Derewenda, Rational protein crystallization by mutational surface engineering. *Structure* **12**, 529-535 (2004).

[123] S. Nauli *et al.*, Polymer-driven crystallization. *Protein Science* **16**, 2542-2551 (2007).

[124] K. S. Prickett, D. C. Amberg, T. P. Hopp, A Calcium-Dependent Antibody for Identification and Purification of Recombinant Proteins. *Biotechniques* **7**, 580-589 (1989).

[125] J. L. Johnson *et al.*, Structural and biophysical characterization of an epitope-specific engineered Fab fragment and complexation with membrane proteins: implications for co-crystallization. *Acta Crystallogr D* **71**, 896-906 (2015).

[126] J. W. Fairman *et al.*, Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic E. coli and enteropathogenic Y. pseudotuberculosis. *Structure* **20**, 1233-1243 (2012).

[127] M. Tomita, K. Tsumoto, Hybridoma technologies for antibody production. *Immunotherapy-Uk* **3**, 371-380 (2011).

[128] D. Rothlisberger, K. M. Pos, A. Pluckthun, An antibody library for stabilizing and crystallizing membrane proteins - selecting binders to the citrate carrier CitS. *Febs Letters* **564**, 340-348 (2004).

[129] H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, A. Pluckthun, Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology* **332**, 489-503 (2003).

[130] P. Milovnik, D. Ferrari, C. A. Sarkar, A. Pluckthun, Selection and characterization of DARPins specific for the neurotensin receptor 1. *Protein Engineering Design & Selection* **22**, 357-366 (2009).

[131] G. E. Dale, C. Oefner, A. D'Arcy, The protein as a variable in protein crystallization. *J Struct Biol* **142**, 88-97 (2003).

[132] H. E. Mcelroy, G. W. Sisson, W. E. Schoettlin, R. M. Aust, J. E. Villafranca, Studies on Engineering Crystallizability by Mutation of Surface Residues of Human Thymidylate Synthase. *J Cryst Growth* **122**, 265-272 (1992).

[133] A. Wernimont, A. Edwards, In situ proteolysis to generate crystals for structure determination: an update. *PLoS One* **4**, e5094 (2009).

[134] D. R. Smyth, M. K. Mrozkiewicz, W. J. McGrath, P. Listwan, B. Kobe, Crystal structures of fusion proteins with large-affinity tags. *Protein Science* **12**, 1313-1322 (2003).

[135] R. Ujwal, J. U. Bowie, Crystallizing membrane proteins using lipidic bicelles. *Methods* **55**, 337-341 (2011).

[136] M. Caffrey, V. Cherezov, Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc* **4**, 706-731 (2009).

[137] N. Cerdà-Costa, F. Xavier Gomis-Rüth, Architecture and function of metallopeptidase catalytic domains. *Protein Science* **23**, 123-144 (2014).

[138] T. S. Thorsen, R. Matt, W. I. Weis, B. K. Kobilka, Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis. *Structure* **22**, 1657-1664 (2014).

[139] J. A. Delmar, J. R. Bolla, C. C. Su, E. W. Yu, Crystallization of membrane proteins by vapor diffusion. *Methods Enzymol* **557**, 363-392 (2015).

[140] D. Veesler, S. Blangy, C. Cambillau, G. Sciara, There is a baby in the bath water: AcrB contamination is a major problem in membrane-protein crystallization. *Acta Crystallogr F* **64**, 880-885 (2008).

[141] G. Psakis, J. Polaczek, B. O. Essen, AcrB et al.: Obstinate contaminants in a picogram scale. One more bottleneck in the membrane protein structure pipeline. *Journal of Structural Biology* **166**, 107-111 (2009).

[142] B. Wiseman *et al.*, Stubborn Contaminants: Influence of Detergents on the Purity of the Multidrug ABC Transporter BmrA. *Plos One* **9**, (2014).

[143] S. Murakami, R. Nakashima, E. Yamashita, A. Yamaguchi, Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587-593 (2002).

[144] S. Faham, J. U. Bowie, Bicelle crystallization A new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *Journal of Molecular Biology* **316**, 1-6 (2002).

[145] C. A. P. Glover *et al.*, AcrB contamination in 2-D crystallization of membrane proteins: Lessons from a sodium channel and a putative monovalent cation/proton antiporter. *Journal of Structural Biology* **176**, 419-424 (2011).

[146] V. Cherezov *et al.*, High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).

[147] M. McGuinness, D. Dowling, Plant-Associated Bacterial Degradation of Toxic Organic Compounds in Soil. *Int J Env Res Pub He* **6**, 2226-2247 (2009).

[148] J. C. Spain, Biodegradation of Nitroaromatic Compounds. *Annu Rev Microbiol* **49**, 523-555 (1995).

[149] Z. C. Symons, N. C. Bruce, Bacterial pathways for degradation of nitroaromatics. *Natural Product Reports* **23**, 845-850 (2006).

[150] K. S. Ju, R. E. Parales, Nitroaromatic compounds, from synthesis to biodegradation. *Microbiol Mol Biol Rev* **74**, 250-272 (2010).

[151] R. Parry, S. Nishino, J. Spain, Naturally-occurring nitro compounds. *Nat Prod Rep* **28**, 152-167 (2011).

[152] R. Winkler, C. Hertweck, Biosynthesis of nitro compounds. *Chembiochem* **8**, 973-977 (2007).

[153] A. Kobayashi, Y. O. Kobayashi, N. Someya, S. Ikeda, Community analysis of root- and tuber-associated bacteria in field-grown potato plants harboring different resistance levels against common scab. *Microbes Environ* **30**, 301-309 (2015).

[154] S. Compant, B. Duffy, J. Nowak, C. Clement, E. A. Barka, Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. *Appl Environ Microbiol* **71**, 4951-4959 (2005).

[155] S. Diallo *et al.*, Mechanisms and recent advances in biological control mediated through the potato rhizosphere. *FEMS Microbiol Ecol* **75**, 351-364 (2011).

[156] R. R. King, C. H. Lawrence, L. A. Calhoun, Unusual production of 5-nitroanthranilic acid by Streptomyces scabies. *Phytochemistry* **49**, 1265-1267 (1998).

[157] R. R. King, L. A. Calhoun, The thaxtomin phytotoxins: sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry* **70**, 833-841 (2009).

[158] S. Lerat, A. M. Simao-Beaunoir, C. Beaulieu, Genetic and physiological determinants of Streptomyces scabies pathogenicity. *Mol Plant Pathol* **10**, 579-585 (2009).

[159] Y. Qu, J. C. Spain, Biodegradation of 5-nitroanthranilic acid by Bradyrhizobium sp. strain JS329. *Appl Environ Microbiol* **76**, 1417-1422 (2010).

[160] P. Prashar, N. Kapoor, S. Sachdeva, Rhizosphere: its structure, bacterial diversity and significance. *Reviews in Environmental Science and Bio/Technology* **13**, 63-77 (2013).

[161] S. G. Conticello, M. A. Langlois, M. S. Neuberger, Insights into DNA deaminases. *Nat Struct Mol Biol* **14**, 7-9 (2007).

[162] A. R. Lambert *et al.*, Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure* **16**, 558-569 (2008).

[163] J. Sambrook, D. W. Russell, *Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., ed. 3rd, 2001), pp. 3 v.

[164] S. D. Orwig, R. L. Lieberman, Biophysical characterization of the olfactomedin domain of myocilin, an extracellular matrix protein implicated in inherited forms of glaucoma. *PLoS One* **6**, e16347 (2011).

[165] S. F. Nishino, J. C. Spain, Biodegradation of 3-nitrotyrosine by Burkholderia sp strain JS165 and Variovorax paradoxus JS171. *Appl Environ Microb* **72**, 1040-1044 (2006).

[166] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307-326 (1997).

[167] T. C. Terwilliger *et al.*, Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D* **65**, 582-601 (2009).

[168] A. J. Mccoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

[169] N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D* **65**, 1074-1080 (2009).

[170] E. Mathew, A. Mirza, N. Menhart, Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. *J Synchrotron Radiat* **11**, 314-318 (2004).

[171] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277-1282 (2003).

[172] D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879-2886 (1999).

[173] V. V. Volkov, D. I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* **36**, 860-864 (2003).

[174] M. B. Kozin, D. I. Svergun, Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33-41 (2001).

[175] H. McWilliam *et al.*, Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597-600 (2013).

[176] J. M. Yang, C. H. Tung, Protein structure database search and evolutionary classification. *Nucleic Acids Research* **34**, 3646-3659 (2006).

[177] C. H. Tung, J. W. Huang, J. M. Yang, Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. *Genome Biology* **8**, 16 (2007).

[178] E. Chovancova *et al.*, CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**, e1002708 (2012).

[179] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31**, 3381-3385 (2003).

[180] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729 (2013).

[181] R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

[182] D. T. Jones, W. R. Taylor, J. M. Thornton, The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275-282 (1992).

[183] S. Martinez-Rodriguez *et al.*, Mutational and structural analysis of L-N-carbamoylase reveals new insights into a peptidase M20/M25/M40 family member. *J Bacteriol* **194**, 5759-5768 (2012).

[184] V. Pelmenschikov, M. R. A. Blomberg, P. E. Siegbahn, A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J Biol Inorg Chem* **7**, 284-298 (2002).

[185] R. S. Hall *et al.*, Three-Dimensional Structure and Catalytic Mechanism of Cytosine Deaminase. *Biochemistry* **50**, 5077-5085 (2011).

[186] R. B. Silverman, *The organic chemistry of enzyme-catalyzed reactions*. (Academic Press, San Diego, ed. Rev., 2002), pp. xvii, 717 p.

[187] A. Slobodkin, A. L. Reysenbach, F. Mayer, J. Wiegel, Isolation and characterization of the homoacetogenic thermophilic bacterium Moorella glycerini sp. nov. *Int J Syst Bacteriol* **47**, 969-974 (1997).

[188] R. N. Armstrong, Glutathione S-transferases: reaction mechanism, structure, and function. *Chem Res Toxicol* **4**, 131-140 (1991).

[189] A. W. Foster, D. Osman, N. J. Robinson, Metal preferences and metallation. *J Biol Chem* **289**, 28095-28103 (2014).

[190] D. P. Barondeau, E. D. Getzoff, Structural insights into protein-metal ion partnerships. *Curr Opin Struct Biol* **14**, 765-774 (2004).

[191] A. Cvetkovic *et al.*, Microbial metalloproteomes are largely uncharacterized. *Nature* **466**, 779-782 (2010).

[192] J. A. Gerlt, P. C. Babbitt, Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* **70**, 209-246 (2001).

[193] V. C. Culotta, M. J. Daly, Manganese complexes: diverse metabolic routes to oxidative stress resistance in prokaryotes and yeast. *Antioxid Redox Signal* **19**, 933-944 (2013).

[194] J. R. Jefferson, J. B. Hunt, A. Ginsburg, Characterization of Indo-1 and Quin-2 as Spectroscopic Probes for Zn-2+-Protein Interactions. *Analytical Biochemistry* **187**, 328-336 (1990).

[195] P. Fromme, J. C. H. Spence, Femtosecond nanocrystallography using X-ray lasers for membrane protein structure determination. *Curr Opin Struc Biol* **21**, 509-516 (2011).

[196] M. S. Hunter *et al.*, X-ray Diffraction from Membrane Protein Nanocrystals. *Biophys J* **100**, 198-206 (2011).

[197] U. Weierstall *et al.*, Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Commun* **5**, (2014).

[198] H. Tanabe *et al.*, Crystal structures of the human adiponectin receptors. *Nature* **520**, 312-+ (2015).

[199] L. Qin *et al.*, Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117-1122 (2015).

[200] G. Fenalti *et al.*, Structural basis for bifunctional peptide recognition at human delta-opioid receptor. *Nat Struct Mol Biol* **22**, 265-268 (2015).

[201] T. Weinert *et al.*, Fast native-SAD phasing for routine macromolecular structure determination. *Nature Methods* **12**, 131-U163 (2015).

[202] G. G. Prive, Lipopeptide detergents for membrane protein studies. *Curr Opin Struc Biol* **19**, 379-385 (2009).

[203] C. L. McGregor *et al.*, Lipopeptide detergents designed for the structural study of membrane proteins. *Nature Biotechnology* **21**, 171-176 (2003).

[204] K. R. Vinothkumar, Membrane protein structures without crystals, by single particle electron cryomicroscopy. *Curr Opin Struc Biol* **33**, 103-114 (2015).

[205] E. Callaway, The Revolution Will Not Be Crystallized. *Nature* **525**, 172-174 (2015).

[206] S. G. Patching, Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. *Bba-Biomembranes* **1838**, 43-55 (2014).

[207] A. L. Brunkan, M. Martinez, J. Wang, E. S. Walker, A. M. Goate, A domain at the C-terminus of PS1 is required for presenilinase and gamma-secretase activities. *J Neurochem* **92**, 1158-1169 (2005).

[208] S. Takagi-Niidome *et al.*, Cooperative roles of hydrophilic loop 1 and the C-terminus of presenilin 1 in the substrate-gating mechanism of gamma-secretase. *J Neurosci* **35**, 2646-2656 (2015).

[209] A. K. Somavarapu, K. P. Kepp, The dynamic mechanism of presenilin-1 function: Sensitive gate dynamics and loop unplugging control protein access. *Neurobiol Dis* **89**, 147-156 (2016).

[210] X. C. Bai, E. Rajendra, G. Yang, Y. Shi, S. H. Scheres, Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, (2015).

[211] A. Tolia, K. Horre, B. De Strooper, Transmembrane domain 9 of presenilin determines the dynamic conformation of the catalytic site of gamma-secretase. *J Biol Chem* **283**, 19793-19803 (2008).

[212] Y. Li *et al.*, Structural interactions between inhibitor and substrate docking sites give insight into mechanisms of human PS1 complexes. *Structure* **22**, 125-135 (2014).

[213] X. C. Bai *et al.*, An atomic structure of human gamma-secretase. *Nature* **525**, 212-217 (2015).

[214] T. Tomita, T. Iwatsubo, Structural biology of presenilins and signal peptide peptidases. *J Biol Chem* **288**, 14673-14680 (2013).

[215] C. Sato, Y. Morohashi, T. Tomita, T. Iwatsubo, Structure of the catalytic pore of gamma-secretase probed by the accessibility of substituted cysteines. *Journal of Neuroscience* **26**, 12081-12088 (2006).

[216] C. Sato, S. Takagi, T. Tomita, T. Iwatsubo, The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the gamma-secretase. *J Neurosci* **28**, 6264-6271 (2008).

[217] Y. Morohashi *et al.*, C-terminal fragment of presenilin is the molecular target of a dipeptidic gamma-secretase-specific inhibitor DAPT (N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester). *Journal of Biological Chemistry* **281**, 14670-14676 (2006).

[218] T. Tomita *et al.*, The first proline of PALP motif at the C terminus of presenilins is obligatory for stabilization, complex formation, and gamma-secretase activities of presenilins. *Journal of Biological Chemistry* **276**, 33273-33281 (2001).

[219] J. Wang, A. L. Brunkan, S. Hecimovic, E. Walker, A. Goate, Conserved "PAL" sequence in presenilins is essential for gamma-secretase activity, but not required for formation or stabilization of gamma-secretase complexes. *Neurobiol Dis* **15**, 654-666 (2004).

[220] J. Wang *et al.*, C-terminal PAL motif of presenilin and presenilin homologues required for normal active site conformation. *J Neurochem* **96**, 218-227 (2006).

[221] M. Cruts, J. Theuns, C. Van Broeckhoven, Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat* **33**, 1340-1344 (2012).

VITA

SIBEL KALYONCU

Sibel was born in Kocaeli, Turkey located 60 miles away from eastern side of Istanbul. She graduated from secondary school as first in her class in 1999. She was admitted to Kocaeli Anatolian High School with her high scores in the national entrance exam. Then, she graduated with two BSc degrees (Molecular Biology & Genetics and Chemistry) from Bogazici University in 2008. She had scholarship from Turkish Education Foundation during her undergraduate. With the help of full scholarship from Koc Foundation, she earned her MSc degree in Chemical and Biological Engineering at Koc University in 2010. In 2011, she moved to Atlanta, GA, US for her PhD in Chemistry and Biochemistry at Georgia Institute of Technology. Her PhD research was funded by grants from the NIH (DK091357 to RLL), NSF (0845445 to RLL), Pew scholars (to RLL) and Georgia Tech Senior Biophysics Training award (to SK).